SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 19	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Robust and Nonparametric Detection of Fading Narrowband Signals	5. TYPE OF REPORT & PERIOD COVERED: Technical Report, Sept. 1983–August 1985
6. PERFORMING ORG. REPORT NUMBER	
7. AUTHOR(s): M. Weiss and Stuart C. Schwartz	8. CONTRACT OR GRANT NUMBER(s): N00014-81-K-0146
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Information Sciences & Systems Laboratory, Dept. of Electrical Eng. & computer Sci., Princeton University, Princeton, NJ 08544	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR SRO-103
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research (Code 411SP), Department of the Navy, Arlington, Virginia 22217	12. REPORT DATE: January 1986 13. NUMBER OF PAGES: 237
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Also submitted as the Ph.D. Thesis of Moshe Weiss to the EECS Department, Princeton University, Princeton, NJ 08544, August 1985.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

fading narrow-band signals
robust detection
nonparametric detection
scale invariance

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This thesis deals primarily with the design and analysis of algorithms for detection of narrow-band slowly fading signals, embedded in noise of uncertain distribution. The signal's phase and amplitude are random, but are random, but are essentially constant over the observation interval; in adition, the Doppler frequency of the signal is unknown. This is the most common model of target echoes received by radar or sonar systems.

Attention is focused on i.i.d. noise samples whose marginal distribution belongs to an ε-mixture family, where the scale (variance) of the "nominal"

Cont.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Continued

part of the distribution is also unknown. This is an important generalization of the model. The purpose of the derived detection algroithms is twofold: first, to assure constant probability of false alarm, regardless of the noise distribution (DF-CFAR); and second, to achieve optimal maximin probability of detection (P_d) performance.

Unavoidably, the performance analysis is asymptotic, but the main objective is in applications involving small sample sizes (n). Therefore, several different tests which are asymptotically equivalent are proposed, and then their small sample behavior is compared by means of Monte-Carlo simulations. In all the algorithms, the test statistic preserves the structure of the quadrature matched filter, which is optimal for signal of unknown phase in Gaussian noise. However, in place of the linear sample mean, a minimax robust estimator of the random amplitude is substituted. To generate this estimate, scale invariant M- (maximum likelihood type), L- (linear combination of order statistics) or R- (rank) type statistics are utilized.

The DF-CFAR property is obtained by using additional noise-reference observations, which are available in a matrix format. One of the columns consists of the samples taken from the spatial resolution cell which presumably contains the target, and the others are taken from adjacent cells. Several canonical structures, denoted as sliding-window (SW), one-sample (1S), two-sample (2S), and single-sweep (SS) are studied (the last two are compatible only with R- type statistics). A new type of 1S R- test is proposed, which does not use the signs of the observations and thus is CFAR even for noise with asymmetrical distributions. (This test requires bi-phase modulations of the transmitted pulse train.) The R- type tests are inherently DF-CFAR, even for finite n. The others achieve it only asymptotically, but simulations have shown rapid convergence.

In addition to differences in complexity and small sample size sensitivity, the tests vary in their performance in a non-homogeneous noise (clutter) environment. The 1S is inherently immune to degradations caused by spatial non-stationarity, and the SW can be protected in a simple manner against the important types. On the other hand, only the new SS test for signal of unknown Doppler frequency preserves the complexity of the conventional periodogram test. In this test, each of the complex observations from the target's cell are ranked with respect to their simultaneous noise reference observations, and then an FFT is performed on the vector of robustly scored ranks. In all the other tests, the robust estimators/statistics must be generated for each of the n unknown frequencies. The behavior of the tests in dependent noise situations is briefly discussed, and modifications are proposed.

The problem of robust digital M-ary communication in white non-Gaussian noise is also studied. With some simplifying assumptions, it is possible to cast it into a framework similar to that of robust estimation of location parameters. The proposed receivers preserve the structure of the MAP decision rule for Gaussian noise, but the M correlators are replaced by limiter-correlators or by generalized M-estimators. Expressions for the error probability are derived for both coherent and noncoherent receivers; they are shown to be identical with those for the AWGN channel, but the SNR is multiplied by the same functional that characterizes M-estimators. Monotonicity of the error probability with respect to this effective SNR ensures minimax performance.

REPORT NUMBER 19

ROBUST AND NONPARAMETRIC DETECTION OF FADING NARROWBAND SIGNALS

M. WEISS and S.C. SCHWARTZ

INFORMATION SCIENCES AND SYSTEMS LABORATORY
Department of Electrical Engineering
Princeton University
Princeton, New Jersey 08544

JANUARY 1986

Prepared for

OFFICE OF NAVAL RESEARCH (Code 411SP)
Statistics and Probability Branch
Arlington, Virginia 22217
under Contract N00014-81-K0146
Program in Ocean Surveillance and Signal Processing

S.C. Schwartz, Principal Investigator

ROBUST AND NONPARAMETRIC DETECTION OF FADING NARROWBAND SIGNALS

Moshe Weiss

ABSTRACT

This thesis deals primarily with the design and analysis of algorithms for detection of narrow-band slowly fading signals, embedded in noise of uncertain distribution. The signal's phase and amplitude are random, but are essentially constant over the observation interval; in addition, the Doppler frequency of the signal is unknown. This is the most common model of target echoes received by radar or sonar systems.

Attention is focused on i.i.d. noise samples whose marginal distribution belongs to an ϵ–mixture family, where the scale (variance) of the "nominal" part of the distribution is also unknown. This is an important generalization of the model. The purpose of the derived detection algorithms is twofold: first, to assure constant probability of false alarm, regardless of the noise distribution (DF-CFAR); and second, to achieve optimal maximin probability of detection (P_d) performance.

Unavoidably, the performance analysis is asymptotic, but the main objective is in applications involving small sample sizes (n). Therefore, several different tests which are asymptotically equivalent are proposed, and then their small sample behavior is compared by means of Monte-Carlo simulations. In all the algorithms, the test statistic preserves the structure of the quadrature matched filter, which is optimal for signal of unknown phase in Gaussian noise. However, in place of the linear sample mean, a minimax robust estimator of the random amplitude is substituted. To generate this estimate, scale invariant M- (maximum likelihood type), L- (linear combination of order statistics) or R- (rank) type statistics are utilized.

The DF-CFAR property is obtained by using additional noise-reference observations, which are available in a matrix format. One of the columns consists of the samples

taken from the spatial resolution cell which presumably contains the target, and the others are taken from adjacent cells. Several canonical structures, denoted as sliding-window (SW), one-sample (1S), two-sample (2S) and single-sweep (SS) are studied (the last two are compatible only with R- type statistics). A new type of 1S R- test is proposed, which does not use the signs of the observations and thus is CFAR even for noise with asymmetrical distributions. (This test requires bi-phase modulations of the transmitted pulse train). The R-type tests are inherently DF-CFAR, even for finite n. The others achieve it only asymptotically, but simulations have shown rapid convergence.

Asymptotic maximin optimal P_d performance is shown for all test structures, with L- and R- statistics, and the simulations demonstrate almost identical performance of the SW- L- type test (with α– trimmed estimators) down to n=16. For the other types, robustness of the P_d against the distribution shape is maintained, but they are somewhat less sensitive. The superiority of the proposed tests over the traditional robust tests, which have been based on a limiter-correlator, is also demonstrated.

In addition to differences in complexity and small sample size sensitivity, the tests vary in their performance in a non-homogeneous noise (clutter) environment. The 1S is inherently immune to degradations caused by spatial non-stationarity, and the SW can be protected in a simple manner against the important types. On the other hand, only the new SS test for signal of unknown Doppler frequency preserves the complexity of the conventional periodogram test. In this test, each of the complex observations from the target's cell are ranked with respect to their simultaneous noise reference observations, and then an FFT is performed on the vector of robustly scored ranks. In all the other tests, the robust estimators/statistics must be generated for each of the n unknown frequencies. The behavior of the tests in dependent noise situations is briefly discussed, and modifications are proposed.

The problem of robust digital M-ary communication in white non-Gaussian noise is also studied. With some simplifying assumptions, it is possible to cast it into a frame-

work similar to that of robust estimation of location parameters. The proposed receivers preserve the structure of the MAP decision rule for Gaussian noise, but the M correlators are replaced by limiter-correlators or by generalized M-estimators. Expressions for the error probability are derived for both coherent and noncoherent receivers; they are shown to be identical with those for the AWGN channel, but the SNR is multiplied by the same functional that characterizes M-estimators. Monotonicity of the error probability with respect to this effective SNR ensures minimax performance.

Table of Contents

1. INTRODUCTION

1.1 Motivation and Outline

The classical theory of signal detection is based on a complete statistical characterization of the signals and interference that are typical of the environment. In many cases, especially for the adverse situations in which radar and sonar systems operate, this precise modeling is either complex or impossible. Consequently, there is a strong motivation to study the design of detection procedures which are insensitive or *robust* to deviations from the assumed statistical models.

Following the fundamental work of Huber on robust estimation of location parameters [1] and on robust hypotheses testing [2], extensive applications and further research have appeared in the communication and information sciences literature, of which references [4]-[12] are a representative sample. Inspection of these references reveals that they are mostly concerned with *lowpass deterministic* (i.e., unknown but non-random) signals, except [9] which treats robust detection of *weak stochastic* signals, and [12] where results have recently been presented for *deterministic narrow-band* signals.

In this thesis, we address the design and analysis of robust detection algorithms which are appropriate for radar - sonar environments. By way of contrast to the deterministic signals that have been studied in the above mentioned references, we focus on coherent, slowly-fading narrow-band signals, which are typical of radar-sonar target echoes. Under this model, the signal's phase and amplitude are random, but are essentially constant over the observation period. Furthermore, the frequency Doppler shift of the received signal is unknown a priori, as it is proportional to the target's radial velocity. While an extensive detection-estimation theory exists for this model when the noise is Gaussian, cf. [13-15], robust detectors have not been investigated. Although the algorithms developed in this thesis share similar techniques for robustification and analysis, the results of [4-12] are not directly applicable. Moreover, in previous studies it has been assumed that the "nominal" noise distribution is precisely known. In this thesis, on the

other hand, we allow uncertainty on its scale, i.e, the noise power. The scale is almost always unknown, as the dominant noise sources are clutter or reverberation echoes, which vary strongly with the environmental conditions. Consequently, one can argue that scale uncertainty is more fundamental than distribution shape uncertainty.

Scale invariant detection has been studied before in the context of constant false alarm rate detection (CFAR); both adaptive and nonparametric techniques have been utilized [21,32]. However, the main objective was to assure the CFAR property and, usually, simultaneous optimization of the probability of detection (P_d) had not been attempted. Apparently, scale invariant detection algorithms which are also insensitive to the distribution shape have not been studied before. (Some tests based on rank statistics are shown in this thesis to be quite robust, but this is a fortunate feature and is not guaranteed a priori).

Another feature of our approach stems from the characteristics of radar - sonar systems. The number of samples (n) is constrained by other system requirements (maximum detection range, spatial resolution, allocated time frame for the detection sector), and typically is between 10 - 50 samples or less. Moreover, the desired levels of the false alarm probability (P_{fa}) are very low; values of 10^{-4} to 10^{-6} are typical. Consequently, optimal algorithms that are the result of an unavoidable asymptotic analysis, serve as candidates for which comparative small-sample simulation studies are performed. While this approach is generally accepted in the statistical literature, only a few of the works [4-12] exhibit small sample performance. Moreover, the majority of the previously devised robust tests are based on *limiter–correlator* structures, where the test statistic is derived from the sum of memoryless transformations on the input observations. For their asymptotic maximin robust optimality, the signal is assumed to be weak. As might be expected, simulation results in this work demonstrate that for small n and low P_{fa}, the limiter-correlator tests suffer considerable P_d losses; clearly, when the signal is sufficiently strong to be detected, it gets limited by the detector.

In this thesis, we look at several different types of tests. Motivated by the good small-sample behavior of robust estimators of a location parameter [35], the proposed tests are based on a statistic which is, essentially, a robust and scale invariant estimator of the amplitude of the narrow-band signal. To generate this estimator, M-, L- or R-type statistics [3] are used. (Tests based on M-estimators were proposed by El-Sawy & VandeLinde [6], Rieder [73], and others for the simpler case of deterministic low-pass signals). Although these different types can be made asymptotically equivalent at a fixed noise distribution by properly choosing their score function, the resulting tests differ in performance when the distribution is allowed to vary over a "neighborhood", and even more so, they differ in their small sample performance. Consequently, by looking at different structures it is possible to choose the more appropriate algorithm for a particular application, in terms of both performance and complexity.

Scale invariance of the amplitude estimator is not sufficient for CFAR operation, and several approaches are studied in this thesis for achieving it. We consider input observations that are available in a matrix format, where one of the columns (of length n) represents the samples taken from the spatial resolution cell which presumably contains the target. The other M columns, taken from adjacent range cells (in the radar or active sonar case), are considered as *noise reference*, i.e., they are assumed to be identically distributed as the target cell is under the null hypothesis. It is possible to utilize the noise reference information in several canonical structures.

In "Sliding Window" (SW) tests, the same statistic is generated for each of the columns; its sample-mean over the reference window (which is an estimator of the variance of the robust amplitude estimator) serves as an adaptive threshold, against which the target cell statistic is compared. The whole structure is then slid over the entire range coordinate, and decisions are made for each of the cells. In this way, the implementation complexity is basically that of computing the single column statistic. All the three types of statistics (M-, L-, R-) are compatible with a SW mechanization.

"Two-sample" tests (2S) are studied in relation to tests based on the ranks of the observations (R-tests). In this class, the decision is based on the ranks of the target's cell observations with respect to the entire reference window observations. In "single-sample" (1S) tests, only the column of the test cell is utilized; these are useful when no reliable assumption concerning the spatial stationarity of the noise can be made. This class is also compatible with all three statistic types. Finally, "single-sweep" (SS) tests are R-type, where for each observation from the test cell, the rank with respect to the same *row* in the matrix is computed, and the final statistic is accumulated over the n rows. This is the only test that preserves the computational efficiency of the linear periodogram (FFT) detector, when implemented as a bank of n frequency contiguous tests over the Doppler uncertainty range.

The tests based on R-statistics enable distribution free (DF) CFAR operation, even for finite n. For the other types DF-CFAR is achieved only asymptotically, but the simulations have shown rapid convergence. In all combinations of the statistic type and reference structure, it is possible to design for an exact or approximate maximin P_d performance: in some class of noise distributions the test is optimal, against a particular distribution, within a specified class of tests; and its P_d is larger for any other distribution in the class. However, the actual P_d depends also on the noise-reference structure. Thus, for a particular application, and for different assumptions on the spatial stationarity, it is possible to trade-off complexity for performance by choosing the appropriate combination of statistic type and reference structure.

As a consequence of the asymptotic analysis and simulation study, our tentative conclusion is that the proposed robust detectors emerge as quite adequate for successfully treating most of the important uncertainties that are encountered in the real-world radar-sonar detection environment.

Outline Of The Thesis

Chapter 2 introduces the problem and presents a performance analysis of the common (unrobustified) quadrature detector under situations of contamination. Since in most applications even if the noise distribution is known its variance is usually not, the adaptive threshold version of the detector is emphasized. Here, the test statistic is normalized by the Maximum-Likelihood (ML) estimate of the noise level. While this test achieves the desired CFAR property asymptotically, it is demonstrated that for small sample sizes P_{fa} can increase intolerably under contamination, even if the percentage of the contaminating component (ϵ) of a mixture distribution [1] is rather small. Both analytical approximation of the distribution (by means of the first terms in a Laguerre type expansion) and simulation are utilized. In addition to P_{fa} degradation, a large decrease in P_d is also possible.

In Chapter 3 we study a very simple robust test on the "coherent envelope" of the observables, i.e., on the output of a quadrature matched filter which is the sufficient statistic under the purely Gaussian noise case. For this scalar statistic, extension of Huber's [2] maximin robust test is straightforward. It consists essentially in finding the threshold setting and a randomization constant under the least favorable condition. However, it is found that the proposed test can protect against contaminations that would have roughly doubled the P_{fa} of the unrobustified quadrature receiver. Beyond that, the maximin bound on P_d decreases rapidly as a function of ϵ. Hence, this test is rather inappropriate for radar/sonar, where very low P_{fa} is required, and ϵ can be much larger than the desired P_{fa}.

In the following chapters, the problem is solved in stages. We first treat the case where only the signal's phase is random; the full problem with all the previously mentioned uncertainties is addressed in chapters 5 and 6.

In Chapter 4 an asymptotically robust test, for detection of narrowband signals of known frequency and additive noise of uncertain distribution with known scale, is

proposed and analyzed. The common large sample assumption is an unavoidable necessity in order to get explicit functionals for the error probabilities, on which optimization according to a maximin criterion becomes possible. However, we abandon the weak signal assumption that has dominated the literature. By avoiding this assumption, we hope that the proposed test will be adequate for small sample sizes. The proposed test statistic is derived from a minimax robust ML estimation of the amplitude. This is achieved by applying Huber's [1] M-estimator of "location" to both the in-phase and quadrature samples, and then independence of the random phase is obtained by summing the squares of these estimates, exactly as in the quadrature matched filter. This test is shown to be maximin robust for sufficiently large values of the desired P_d (in the range of practical interest) and for several common models of the target's amplitude distribution. Finally, the weak signal locally optimal robust detector for our problem is outlined, and is shown to be a local approximation to the M-estimation/detection structure derived here.

In Chapter 5, further uncertainties that must be considered in practical applications are studied. With regard to ϵ in the mixture model, it is shown numerically that even a very pessimistic design with ϵ=0.5 incurs a rather small additional loss compared to the case when ϵ is known exactly. Thus, exact knowledge of ϵ is unnecessary. More important, the robust M-detector is extended to handle an unknown scale (power level) of the nominal noise distribution in the mixture family, by coupling it with a robust estimator of the scale, from which an adaptive threshold is generated. While it is not possible to exhibit maximin properties of the extended test in a 1S class (i.e., when no reference observations are available), it is shown numerically to be qualitatively robust in the sense that an upper bound on P_{fa} and a lower bound on P_d can be guaranteed within various parametrized mixture families. The latter bound is quite close to the P_d obtained by the optimal detector for the nominal Gaussian noise.

When noise reference is available, a similar estimator-detector with an adaptive

threshold, which serves as an estimator of the variance of the amplitude estimator, is shown to be maximin in the class of SW detectors. Due to its CFAR property for *any* noise distribution, the restriction for sufficiently high P_d of Chapter 4 is relaxed. Moreover, the maximin performance is valid for arbitrary distribution of the signal's amplitude, and when the reference window size M is also large, the test becomes globally maximin optimal (i.e., without restriction to the SW class).

For the important case of nominal Gaussian noise, the M-estimators in either a SW or a 1S test can be replaced by α– trimmed estimators, which are asymptotically equivalent in probability and simpler to compute. The same structure can be generalized for non-Gaussian *nominal* p.d.f. with the appropriate optimal robust L-estimators. Finally, the signal frequency which is also unknown in any realistic application (due to Doppler shifts), is treated by constructing a bank of contiguous robust tests which covers the uncertainty range. It is shown that the detectability loss for signals whose frequencies straddle between adjacent such "filters", is asymptotically identical with that incurred by the usual periodogram (FFT) detector.

The price that is paid for the improved capability with these new receiver structures is a substantial increase in signal processing complexity, compared to the usual linear FFT test. However, the most demanding nonlinear processing task required here consists of several levels of data rankings, and this operation is becoming possible even for real-time radar applications with VLSI and VHSIC technology [16].

The final section of Chapter 5 presents a thorough Monte-Carlo study of the small-sample performance of the various robust detectors. The simulation results show that the performance predicted by the asymptotic analysis of the SW robust detectors is essentially maintained even for sample sizes as small as $n=16$, in guaranteeing a high lower bound on the detection probability as well as in controlling the false alarm-level. For the 1S tests, somewhat larger sample sizes are necessary for convergence to the asymptotic prediction. By way of contrast, the corresponding weak-signal locally

optimal robust detector is shown to produce high losses for small desired P_{fa} and large deviation of the signal frequency. These losses are attributed to poor convergence properties to the Gaussian distribution as a result of hard limiting on the test statistic, especially when the signal is sufficiently large to allow high detection probability. On the other hand, our proposed tests limit the influence of outliers *around* the arbitrary signal amplitude, and they converge rapidly to the asymptotic values.

Chapter 6 treats detection of narrowband signals by rank based statistics. The motivation is twofold - first, to achieve DF-CFAR even with finite sample sizes; and second, to improve the efficiency of the SW and 1S tests based on robust estimators. For a noise distribution which is known in shape, asymptotically optimal tests in all three generic classes (1S, 2S and SS) are found. Of prime importance is the new type in the 1S class. It is shown that if the polarity of the transmitted signal is changed in half the samples, it is possible to obtain with a 1S rank test optimal P_d performance, identical to that of the completely known distribution case, and without utilizing any reference samples. Identical performance is possible with a SW test only for large M. Consequently, the 1S R-test is inherently robust in a spatially non-stationary environment.

In the SS class, a Doppler test that requires the least amount of computations is found. First, the target cell observations are ranked with respect to the same row, and then an FFT is performed on the complex vector of the scores of the ranks. Thus, the complexity is roughly $O(n \log n)$, compared to n times more in Doppler tests with SW, 1S or 2S structures. This test is attractive, therefore, in typical radar systems where even the FFT mechanization requires rates on the order of millions of operations per second.

For all the rank based tests, robustification against the noise distribution uncertainty is achieved by designing the score function according to the least favorable noise distribution. We show, however, that the rank tests with the traditional Wilcoxon or Van Der Waerden score functions are inherently robust. Nevertheless, the robust score

function does not entail significant additional computation complexity, and should be preferred.

The chapter ends with a Monte-Carlo study of the small sample behavior, followed by a comparison between the different robust tests of previous chapters. The issues of compatibility with a non-homogeneous environment, and with dependent noise situations are briefly discussed. It is not appropriate to declare one of the tests as the uniformly best, but it is possible to choose the one more suitable under specific environmental conditions, and for an accepted complexity in implementation.

The problem of robust digital M-ary communication receivers is studied in Chapter 7. With some simplifying assumptions, it is possible to cast the Bayesian multiple hypothesis decision problem in a framework similar to that of the binary Neyman-Pearson case. Robustification against non-Gaussian noise is achieved by replacing the linear correlators, in the bank of correlators receiver which is optimal for Gaussian noise, with limiter-correlators or with regression type M-estimators. Implicit expressions for the error probability with white, but non-Gaussian, noise are derived assuming discrete time processing. These expressions are identical in form with those for the AWGN channel, but the effective SNR is now multiplied by the same functional of the M-estimators (both coherent and noncoherent receivers are considered). Consequently, choosing Huber's function for the limiter nonlinearity again results in minimax performance. It is also demonstrated that the capacity of the assumed non-Gaussian channel can be much higher when compared to the AWGN channel of equal noise and signal powers. Therefore, efficient and robust communication is possible with either higher rates or reduced transmitted power.

1.2 Review of Huber's Results for Maximin Testing of Simple Hypotheses

Huber [2] considered and solved the following problem. Let $\{x_i\}_{i=1}^n$ be a sequence of independent random variables, and let $\{P_0,P_1\}$ be distinct probability measures on the real line with the corresponding densities $\{f_0,f_1\}$ with respect to some measure. Assume that the likelihood ratio (LR) $f_1(x)/f_0(x) = L(x)$ almost surely, where L(x) is a monotone function.

Let **M** be the set of all probability measures on the real line and $0<\epsilon<1$ a given number. The uncertainty in the distribution of the observations is introduced by expanding the *simple* hypothesis P_0 and *simple* alternative P_1 into composite ones by a mixture model-

$$H_0: \quad \mathbf{P}_0 = \{Q \in \mathbf{M} \mid Q = (1-\epsilon)P_0 + \epsilon C_0, \quad C_0 \in \mathbf{M}\} \tag{1.1a}$$

$$H_1: \quad \mathbf{P}_1 = \{Q \in \mathbf{M} \mid Q = (1-\epsilon)P_1 + \epsilon C_1, \quad C_1 \in \mathbf{M}\} \tag{1.1b}$$

Actually, Huber's setup is more general, it allows also the following neighborhoods of the nominal model: total variation, Prohorov distance, Kolmogorov distance and Lévy distance. Moreover, different ϵ_i are allowed in the H_i.

The problem is to find the most robust test in the maximin sense between $\mathbf{P}_0$ and $\mathbf{P}_1$, i.e., to find a saddle-point pair of test $d^*(\mathbf{x}) \in \mathbf{D}$, where **D** is the class of all decision rules, and distributions $q_i^* \in \mathbf{M}$ such that

$$\sup_{d \in \mathbf{D}} \beta(d,q_1^*) \;=\; \beta(d^*,q_1^*) \;=\; \inf_{q \in \mathbf{P}_1} \beta(d^*,q) \;, \tag{1.2}$$

subject to

$$\sup_{\substack{d \in \mathbf{D} \\ q \in \mathbf{P}_0}} \alpha(d,q) \;=\; \alpha(d^*,q_0^*) \;=\; \alpha_0 \tag{1.3}$$

Here, $\beta(d,q)$ is the power of the test (detection probability) at density q:

$$\beta(d,q) \;=\; Prob\,\{d(\mathbf{x})=H_1 \mid H_1 \; is \; true \;\; (q \in \mathbf{P}_1)\} \tag{1.4}$$

and $\alpha(d,q)$ is the level (false alarm probability)

$$\alpha(d,q) \;=\; Prob\,\{d(\mathbf{x})=H_1 \mid H_0 \; is \; true \;\; (q \in \mathbf{P}_0)\} \tag{1.5}$$

The meaning of the criterion is clear- d^* is the best Neyman-Pearson (NP) decision rule for the least-favorable pair $\{q_0^*, q_1^*\}$.

As Huber showed, the most robust test is a NP test on the pair

$$q_0^* = \begin{cases} (1-\epsilon) f_0(x) & L(x) < L'' \\ \frac{(1-\epsilon)}{L''} f_1(x) & L(x) \geq L'' \end{cases} \tag{1.6}$$

$$q_1^* = \begin{cases} (1-\epsilon) f_1(x) & L(x) \geq L' \\ (1-\epsilon) L' f_0(x) & L(x) < L' \end{cases} \tag{1.7}$$

The numbers L' and L'' are determined such that q_0^* and q_1^* are legitimate density functions.

For ϵ sufficiently small (a condition which is equivalent to disjointness of $\mathbf{P}_0$ and $\mathbf{P}_1$) the normalizing equations have a unique solution with $0 \leq L' \leq L'' < \infty$. The LR between q_0^* and q_1^* is thus given by "soft-limiting" the nominal LR to-

$$l(x; L', L'') = \begin{cases} L' & \text{when} \quad L(x) \leq L' \\ L(x) & \text{when} \quad L' < L(x) < L'' \\ L'' & \text{when} \quad L(x) \geq L'' \end{cases} \tag{1.8}$$

and a maximin robust test is a randomized NP test on $T(\mathbf{x}) = \prod_{i=1}^{n} l(x_i; L', L'')$:

$$d^*(\mathbf{x}) = \begin{cases} H_1 & \text{for} \quad T(\mathbf{x}) > t \\ H_1 \text{ with probability } c & \text{for} \quad T(\mathbf{x}) = t \\ H_0 & \text{for} \quad T(\mathbf{x}) < t \end{cases} \tag{1.9}$$

The quantities t and c are determined from the right side equality in Eq.(1.3). Note that in general the test must be randomized for arbitrary α_0 since $T(\mathbf{x})$ takes its values at the limiter end-points with finite probability.

It is interesting to point to some peculiarity in the result. The least-favorable densities are constrained by (1.6-1.7), and in particular the contamination C_1 can not be related in any other way to C_0. While this is reasonable in a game situation against an intelligent opponent, it seems unlikely for signal detection problems where the uncertainty in $\mathbf{P}_1$ is induced by that in $\mathbf{P}_0$ and is not affected by the presence or absence of an

additive signal whose characteristics are assumed to be known. For example, in a "location shift" problem ($H_0: x_i = n_i$, $H_1: x_i = n_i + a$) we would like $\mathbf{P}_1$ to be the class of all distributions that are shifted to the right by a from those in $\mathbf{P}_0$, and specifically $q_1^*(x) = q_0^*(x-a)$, which *can not* also satisfy (1.6-1.7). Thus, it is suggested that a better solution might exist for this physical formulation of the problem; unfortunately, it is unknown [1]. The asymptotic reformulation in Chap. 4 essentially avoids this peculiarity.

It should be emphasized that Huber's proof of optimum robustness relies heavily on two assumptions which are not valid for the problems that are considered in this work: a) H_0 and H_1 are simple hypotheses- P_0 and P_1 do not include any unknown (relevant or nuisance) parameters. b)the observations are independent r.v.'s.

1.3 Review and Critique of Robust Tests Based on Weak Signal Assumption

Often in detection problems, the true value of the signal amplitude A is small but otherwise unknown. In such cases it is plausible, mostly as a theoretical nicety which permits an analytical (and usually quite simple for implementation) solution, to employ the locally optimum detector (LOD) structure. This is the detector structure which maximizes the derivative of the power function (detection probability) at A=0, for a given test level (probability of false-alarm), [25]. Under suitable regularity conditions on the densities, the LOD is identical with the detector which is obtained by taking the leading term of a series expansion for the likelihood-ratio in powers of the SNR around zero [26]. It also maximizes Pitman's efficacy [27] which is a suitable weak signal measure of performance, when the number of the observed samples $n \to \infty$.

The weak signal local optimality criterion was first extended to problems of robust detection by Martin and Schwartz [4], and has been widely applied since then, cf. [5],[7]-[10],[12]. Instead of seeking a maximin relation on the detection probability as in Eq.

[1] Private communication with P.J.Huber,12/1983.

(1.2), it was proposed in [4] to design for a maximin relation on the slope of the asymptotic power function at $A=0$:

$$\sup_{d\in D}\beta'(d,q_0^*) \;=\; \beta'(d^*,q_0^*) \;=\; \inf_{q_0\in \mathbf{P}_0}\beta'(d^*,q_0) \tag{1.10}$$

where $\beta'(d,q)=\frac{d}{dA}\beta(d,q\mid A)\mid_{A=0}$ and subject to the false-alarm constraint of Eq. (1.3). The interpretation is that d^* is the LOD for the least-favorable density q_0^*. This criterion has resulted in all the above mentioned references in a *limiter–correlator* structure: the locally optimal non-linearity of the LOD receiver is robustified against ϵ - contamination by inserting a soft limiter at its output and then correlating the non-linearly transformed observations with the known signal sequence. This will be denoted in the following as LORD - LO Robust Detector.

The performance of LOD schemes is commonly evaluated using asymptotic measures such as the Asymptotic Relative Efficiency (ARE), which requires both the assumptions of large n (to satisfy the central limit theorem) and vanishingly weak SNR. In any practical engineering application, the number of samples must be finite [1] and thus the input SNR must be reasonably large to obtain meaningful detection probability. It follows that LOD schemes might perform poorly in practical applications. Several examples of peculiar and very slow convergence of the finite sample RE to the ARE, can be found in [28]-[31]. These references include examples where the finite-sample/large-signal ranking of detector performance are actually different from ARE prediction. Another new example is given in Appendix A. The large SNR performance of the LOD would be extremely poor when the test non-linearity redescends or even vanishes except for some region around the origin, i.e. $l(x)=0\ \forall\mid x\mid \geq c$. In these cases we will obviously get

[1] In radar-sonar systems, n is directly related to the total search time of the desired sector, to the desired maximum non-ambiguous detection range, and to the spatial resolution of targets. Moreover, the detectability of coherent signals is (asymptotically) governed by the average integrated $SNR=nA^2/2\sigma^2$. Since transmitters are usually constrained by the average rather than peak power, it is only the product nA^2 that matters and improved detectability can be obtained by increasing A as well as n. Hence, in view of the other system design goals that were mentioned in the beginning of this note, n is usually in the range 1-100, and even in 1-3 samples for very long range systems, in contrast with the common theoretical assumptions.

$\lim_{A>>c} \beta \to 0$, in contrast to the desired consistency of the power function with increasing SNR. This situation was actually obtained in [10]. In principle, this undesired property of LOD designs might be corrected by switching between two detectors, where one of them is the LOD and the other one is some amplitude-consistent detector. The switching should occur as a result of a threshold crossing by an estimator of the signal amplitude $\hat{A}$. This estimator must obviously be robust against deviations in the assumed noise model to prevent incorrect switching. While this heuristic proposal has not been analyzed , it does suggest that an optimal robust detector should be based on a robust amplitude estimator, as will be studied in Chap. 4.

Inherently, the LORD scheme is subject to the same consistency problem. In addition, investigation of the previously mentioned references reveals that they all exhibit one or more of the following peculiarities and shortcomings:

i) The support of the contaminating density c in Eq. (1.1) is restricted to lie exclusively on the exterior of some interval $[-x_1,x_1]$, cf. [9]. This might not be a severe limitation since the greatest deterioration in performance of the conventional detectors occurs for contaminations that are in the far tails of the nominal density.

ii) The nominal density itself *is not* a member of the mixture family P_0 in which a least-favorable density is sought, but rather "close" to it in some measure, cf. [9], [12].

iii) The test is maximin robust only for $\alpha \geq \alpha(\epsilon)$, which is roughly in the range 0.05-0.2, cf. [4],[5], and [7]. Hence, no robustness is guaranteed for the most important range of small α: 10^{-4}– 10^{-8}.

It is clear that the maximin relation (1.10) is neither sufficient nor necessary for obtaining the desired maximin solution in terms of the detection probability itself. Since Eqs. (1.2) and (1.3) are contradictory in nature when A can take any value in some interval, the most that can be expected from the LORD approach is as shown in Fig. 1.1: the

shaded region on the β axis indicates the α's obtained for the optimal detector for any density q_0 in the mixture family; the slope of any $\beta(q_0)$ at A=0 agrees with (1.10) and (1.2) is satisfied for all signal amplitudes that are greater than some critical value : $A \geq A_c$ or equivalently $\beta \geq \beta_c$.

However, the power function of a LORD could as well behave as shown in Fig. 1.2 where (1.3) and (1.10) are satisfied, but the curve of $\beta(q_0^*)$ dominates that of $\beta(q_0)$ $\forall$ $A \geq 0$ and $\forall q_0 \in \mathbf{P}_0$. (Other situations are also possible). Nevertheless, if $\forall q_0 \in \mathbf{P}_0$ the distance between the curves in Fig. 1.2 could be made sufficiently small, i.e.- $|\beta(d^*, q_0) - \beta(d^*, q_0^*)| \leq \delta(\epsilon, \alpha_0, A) \rightarrow 0$, the detector would be practically robust, though not in the most general maximin sense. (In that case, q_0^* actually becomes the most-favorable density for the detection probability !)

Closer study of the structure of all the LORD tests in the above mentioned references reveal that implicit in all of them is a "robust estimator of zero amplitude" -i.e., the test is actually based on an approximation of the leading term in the expansion of an appropriate version of Huber's robust estimator in a power series around zero amplitude. This observation establishes another motivation for the asymptotically robust test of Chap. 4 which utilizes robust estimation of the amplitude without any small signal approximation, and thus essentially avoids all of the above mentioned problems.

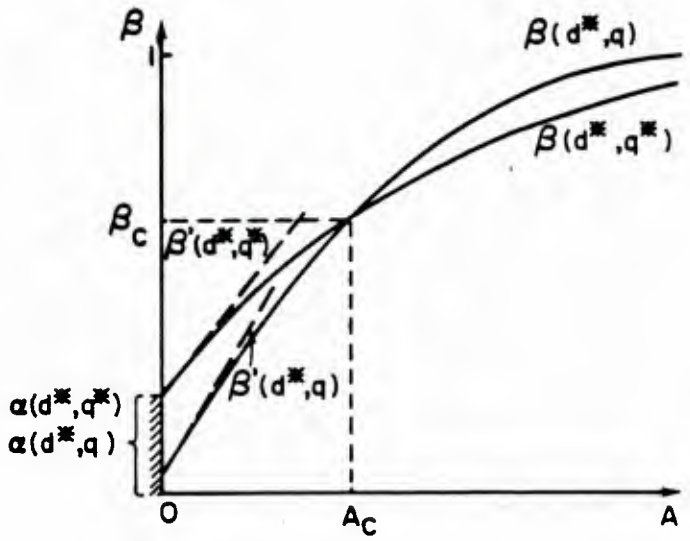


Fig. 1.1 Qualitative power function curves for a desired maximin robust detector of signal with unknown amplitude.

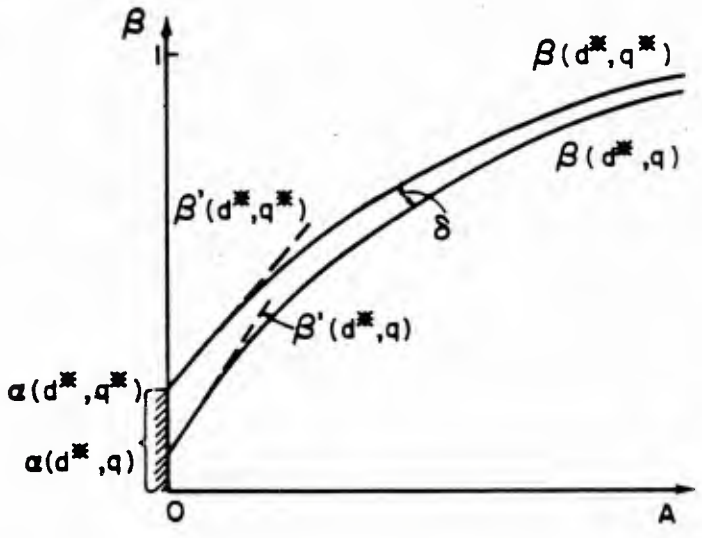


Fig. 1.2 Qualitative power function curves for a locally-optimal robust detector which is not maximin with respect to the detection probability.

2. THE NARROW-BAND SIGNAL AND NOISE DETECTION ENVIRONMENT

2.1 Models and Problem Statement

The problem to be considered is that of robust detection of a slowly fading narrow-band signal with unknown phase and amplitude in nearly Gaussian narrow-band noise. This signal model is typical of most coherent radar targets [13] and also appears in some sonar problems and also in communication over tropospheric links [14].

Specifically, the discrete signal plus noise samples that are observed at the systems input terminals are

$$z_i = As_i \cos(\omega t_i + \phi) + n_i \qquad ,i = 1,...,n \tag{2.1}$$

The n_i are narrow-band noise samples around the *known*[1] center frequency ω :

$$n_i = n_{c_i} \cos\omega t - n_{s_i} \sin\omega t \tag{2.2}$$

$\{\mathbf{n_c},\mathbf{n_s}\}$ are assumed to be zero mean, independent [2] and identically distributed (i.i.d) random variables (r.v.), from a nominally Gaussian ϵ -contaminated mixture :

$$f\,(\mathbf{n_c},\mathbf{n_s}) = \prod_{i=1}^{n} f_{\,0}(n_{c_i}) f_{\,0}(n_{s_i}) \tag{2.3}$$

where

$$f_{\,0}(n_{c_i} = x\,) = f_{\,0}(n_{s_i} = x\,) = (1-\epsilon)\frac{1}{\sqrt{2\pi}\sigma}\exp(\frac{-x^2}{2\sigma^2}) + \epsilon c\,(x\,) \qquad , c \in \mathbf{M} \tag{2.3a}$$

ϕ is the unknown signal phase which is common for all samples (coherent detection) and uniformly distributed over $[0,2\pi]$. It does not convey any information on the target and is thus a nuisance parameter which has to be averaged out. $\{s_i\}_{i=1}^{n}$ is a sequence of known, positive, amplitude modulations [3] with $(1/n)\sum {s_i}^2 = 1$. The amplitude A which

[1] The case of unknown Doppler shift will be treated later by the usual "bank of filters" approach.

[2] Notice that sample -to -sample independence either excludes the common situation of detection in correlated clutter, or it corresponds to a sub-optimal scheme where the observables z_i are obtained by pre- whitening the input sequence.

[3] Known phase modulations can be introduced without much complication.

is also constant for all the observed samples will be considered in the sequel as either deterministic but unknown or random with some known distribution of unknown scale[4].

It is well known, cf. [13], that in the purely Gaussian noise case ($\epsilon = 0$), a uniformly most powerful (UMP) test, independent of the amplitude being deterministic or random with an arbitrary distribution, is equivalent to a threshold test on the "coherent sample envelope" $R(\mathbf{x},\mathbf{y})$:

$$R(\mathbf{x},\mathbf{y}) = (\sum_{i=1}^{n} x_i\, s_i)^2 + (\sum_{i=1}^{n} y_i\, s_i)^2 \tag{2.4}$$

where the in-phase and quadrature samples are obtained by applying the input to a pair of ideal lowpass mixers:

$$x_i = 2z_i \cos\omega t \mid_{LPF} = As_i \cos\phi + n_{c_i} \tag{2.5}$$

$$y_i = 2z_i \sin\omega t \mid_{LPF} = As_i \sin\phi + n_{s_i} \tag{2.6}$$

Since the quadrature components are obtained by a reversible operation, they are equivalent statistics for the problem, which can thus be reformulated as:

$$H_o: \quad \begin{cases} x_i = n_{c_i} \\ y_i = n_{s_i} \end{cases} \tag{2.7a}$$

$$H_1: \quad \begin{cases} x_i = n_{c_i} + As_i \cos\phi \\ y_i = n_{s_i} + As_i \sin\phi \end{cases} \tag{2.7b}$$

It is desired to obtain a maximin relation like Eq. (1.2) for all $A \geq A_c$ (recall the discussion in Section 1.3) subject to the false-alarm constraint Eq. (1.3).

Direct application of Huber's [1] solution for this problem is inappropriate since the samples under H_1 are actually dependent due to the common random phase ϕ, even when A is deterministic and known. However, a "naive" extension of Huber's test is pos-

[4] These slow fading assumptions correspond to common radar targets whose cross section (RCS) fluctuates with a correlation time much longer than the "blip" duration nT, T= intersample period, but much shorter than the scan-to- scan period. This is a most frequent situation [13].

sible if one considers ϕ as a non-random but unknown and adopts an "estimation-detection" procedure. A set of $M \rightarrow \infty$ tests is performed, where the $j^{\underline{th}}$ test is matched to $\phi_j = j/2\pi M$ and its objective is a decision about the presence of a signal *of phase* ϕ_j. Each of the individual tests is then of the *correlator–limiter* type found in [4], but the limiter break-points are now functions of the s_i and also of ϕ_j. Since phase information is irrelevant, the final decision can be reached by accepting H_1 if at least one of the individual tests has accepted it.

We have not studied how many parallel tests have to be performed to achieve acceptable performance with arbitrary signal phases between the adjacent "filters". It is conjectured that it might be prohibitive, especially when the Doppler shift is also unknown so that a two-dimensional "filter-bank" (phase, frequency) has to be constructed.

In the absence of any optimal procedure of design for the finite sample problem, we suggest to weaken somewhat the requirements. We will do that in Chap. 3 by assuming that the observable is only the envelope statistic of Eq. (2.4) rather than the original quadrature samples. We can then construct a most robust test against the uncertainty in the distribution of that *scalar* statistic.

In Chap. 4, we will extend the asymptotic estimation-detection approach of El-Sawy and VandeLinde [6] to our problem, utilizing some of Huber's [1] results on robust estimation of a location parameter. It is shown that if A is replaced by a robust minimax estimate, the resulting test statistic has a limited robust property.

2.2 Performance Degradation of the Conventional Envelope Detector.

In this section, the degradation in performance of the envelope detector of Eq. (2.4), (which is the UMP detector for slow fading narrow-band signals in narrow-band Gaussian noise) will be analyzed when it operates in a background which comes from a Huber-Tukey mixture family. The decrease in detection probability as well as the increase in false alarm probability are of interest, where the latter is of an extreme

importance in an automated search system. These results will serve as a basis for comparison with the performance of the robust detectors that are developed later in this work.

Radar and sonar systems usually operate in a highly non-stationary and non-homogeneous noise background. Except for the receiver thermal noise which is stationary and Gaussian, the statistical properties of interference that is the outcome of external sources, (such as clutter reflections from ground sea or aerial objects in the radar environment, or ambient noise and reverberations in the sonar case [14], [22]), rapidly change in time and space. Therefore, even when the Gaussian assumption is adequate, the fixed threshold detector based on the envelope statistics of Eq. (2.4) is of little value, since its actual false-alarm rate will fluctuate intolerably according to the changes in the background noise level. A common adaptive detector for these situations, which is invariant to the power level of the background noise, compares the envelope statistics from the "test-cell" with an adaptive threshold derived from a "noise-reference" channel. Specifically, $\{\mathbf{x},\mathbf{y}\}$ represent the narrow-band observation samples in the hypothesis-testing cell, $\{\mathbf{u},\mathbf{v}\}$ are the noise-reference samples which are assumed to have *the same distribution* as that of $\{\mathbf{x},\mathbf{y}\}$ under H_0. The adaptive test is:

$$d(\mathbf{x},\mathbf{y},\mathbf{u}.\mathbf{v})=H_1 \qquad if\ f \qquad R(\mathbf{x},\mathbf{y})\geq tW(\mathbf{u},\mathbf{v}) \tag{2.8}$$

where

$$R=I^2+Q^2\ ; \quad I(\mathbf{x})=\frac{1}{n}\sum_{i=1}^{n}x_i\ ; \quad Q(\mathbf{y})=\frac{1}{n}\sum_{1}^{n}y_i \tag{2.9}$$

$$W(\mathbf{u},\mathbf{v})=\frac{1}{M}\sum_{j=1}^{M}(U_j^2+V_j^2) \tag{2.10}$$

$$U_j=\frac{1}{n}\sum_{i=1}^{n}u_{ij} \quad ;V_j=\frac{1}{n}\sum_{i=1}^{n}v_{ij} \tag{2.11}$$

The threshold multiplier t is determined as to achieve the desired false -alarm rate. (Notice that there are M reference vectors for the test vector, all of dimension n).

In search systems where a target presence is to be detected in some spatial sector, the noise reference samples are easily obtainable by applying the same signal processing

of the hypothesis testing channel to adjacent resolution cells in range, Doppler or bearing coordinates, c.f. [18]-[20]. This detector is known as the Cell- Averaging CFAR(CA-CFAR), Mean Level Detector(MLD), or Sliding Window Detector (SW). When the number M of reference samples grow asymptotically, its performance approaches that of the UMP detector for the fixed-variance Gaussian noise. Its structure is also almost identical with that of a narrow-band version which can be derived from the Two-sample Student's-t test [21], which has some optimum properties among unbiased and invariant tests for detection in Gaussian noise of unknown level. (In the extended t-test, the noise variance is estimated from both the test and reference samples in contrast to (2.8). However, the difference in performance is quite small when M is sufficiently large, and the structure of Eq. (2.8) enables SW mechanization where R as well as $U_j^2+V_j^2$ are generated sequentially *by the same hardware* during the search).

In the following we assume that the background interference has an epsilon-contamination mixture density, where the nominal as well as the contaminating densities in Eq.(2.3) are Gaussian: $f = (1-\epsilon)N(0,1)+\epsilon N(0,c^2)$. The performance of the detector is discussed first when the number of samples n is asymptotically large, and secondly for the finite sample case.

a) n $\rightarrow\infty$

When n is asymptotically large, all the quantities I, Q, U_j, and V_j of Eqs.(2.9)-(2.11) are Gaussian random variables. Hence, it is obvious that this detector is asymptotically nonparametric under the null hypothesis; i.e., it is CFAR for any probability density of the input test and reference samples for which the central limit theorem holds.

Under the alternative, for a Rayleigh distributed slow-fading signal the detection-probability of this detector is uniquely determined by the integrated Signal-to-Noise ratio (c.f. [18], [20]): $SNR = nA^2/2\sigma^2$. Since $\sigma^2=(1-\epsilon)+\epsilon c^2$, it is clear that even a small amount of high power(c^2) contamination can reduce the effective SNR by orders of magnitude, thus spoiling detectability completely.

b) Finite n

With regard to the detection probability a similar behavior as above is to be expected, as it is not significantly affected by the tails of the distribution. However, it is no longer true that the detector remains CFAR for moderate and small sample sizes. Since in practice radar-sonar systems are designed for very low false alarm probabilities, it is the tails of the distribution of the test statistics that counts, and even a slight deviation from Gaussianity of the input samples will suffice to substantially increase the area under the tails.

In Appendix B, an approximation is derived for the distribution of the coherent envelope R, which approaches the desired one-sided exponential distribution for large n. [1] The approximation is based on an expansion of the distribution in a series of orthonormal Laguerre polynomials, where its leading term is (with proper normalization such that E(R)=1) the Gamma density

$$f(R)=\frac{\alpha^{\alpha}e^{-\alpha R}R^{\alpha-1}}{\Gamma(\alpha)} \quad , \quad R\geq 0 \tag{2.12}$$

where $\alpha=(1+k/2n)^{-1}$ and k is the kurtosis of the input samples. This is given for a Gauss-Gauss mixture by

$$k = \frac{E(x^4)}{E^2(x^2)} - 3 = 3\epsilon(1-\epsilon)\left[\frac{c^2-1}{1-\epsilon+\epsilon c^2}\right]^2 \tag{2.13}$$

Notice that when $\epsilon \to 0$ but $\epsilon c^2 >> 1$, $k \to 3/\epsilon >> 1$. Table 2.1 demonstrates that k can be very far away from zero for a Gauss-Gauss mixture, thus α is also much smaller than 1 leading to higher false-alarm probability.

[1] Actually, an exact analytic expression was derived for the special case of a Gauss-Gauss mixture, but it was found to be very difficult to compute numerically.

ϵ \ c	3.162	10	31.62
0.001	.238/.992	24.3/.568	748.5/.041
0.01	2.02/.941	73.5/.303	245.4/.115
0.1	6.06/.841	22.3/.589	26.47/.547

Table 2.1 Kurtosis (left entry) and α (right), Gauss-Gauss mixture, n =16.

Figures 2.1 thru 2.5 depicts the false alarm probability vs. the threshold multiplier t, where t is set according to the Gaussian assumption. In each figure, the graphs marked "known sigma" corresponds to the adaptive threshold detector of Eq.(2.8) with M $\rightarrow\infty$; in that case the variance estimation is error-free and the detector is identical in performance to the UMP detector for the known variance case. The graphs marked " ϵ=0" give P_{fa} when the input is indeed Gaussian. In all figures, n=16 and M=8. This is sufficiently small for poor convergence to Gaussianity of the sample means in the tails, but is representative of the number of samples actually employed in most systems. The adaptive-threshold curve with ϵ=0 was computed from $P_{fa}=(1+t/M)^{-M}$, c.f. [18], [20]. The Laguerre approximation was computed only for the M $\rightarrow\infty$ case, since for the adaptive threshold test no similar approximation is possible. [2] The squares were obtained from a Monte-Carlo simulation; it is observed from the figures that the Laguerre approximation is quite reasonable and agrees closely with the simulation results in its region of low variance (roughly down to the reciprocal of the number of runs in the simulation). The main conclusion from the figures is that the increase in P_{fa} is very high, for desired values of 10^{-4} and lower whenever $\alpha<0.9$, and that this deterioration is ordered in correspondence with the deviation of k and α from their values at the Gaussian density. This increase in P_{fa} is much more severe for the adaptive-threshold detector with finite number of reference cells; in Fig. 2.5, for example, the threshold setting for 10^{-6} at the Gaussian produces about $3\cdot10^{-2}$ at the contaminated mixture.

[2] Since it is not possible to obtain analytic expressions for the moments of the ratio R/W of Eq. (2.8). While it is possible to derive a similar approximation to the distribution of W and then to integrate numerically Pr{R>tW}, this approach was not taken.

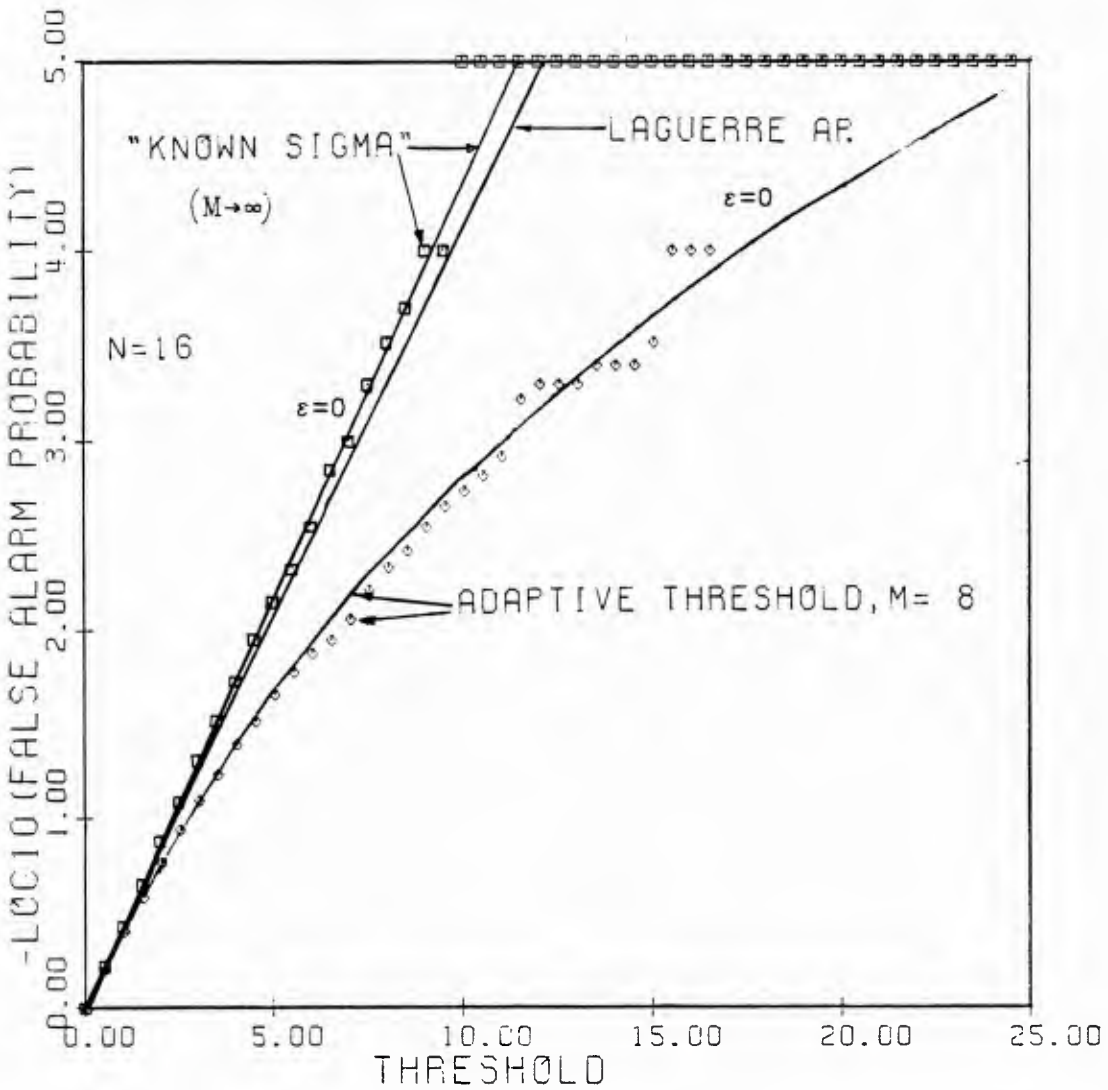


Fig. 2.1 false-alarm probability vs. threshold multiplier. Normal-Normal mixture, ε=0.001,c=3.162,k=.2384,α=.9926,10,000 monte-carlo runs.

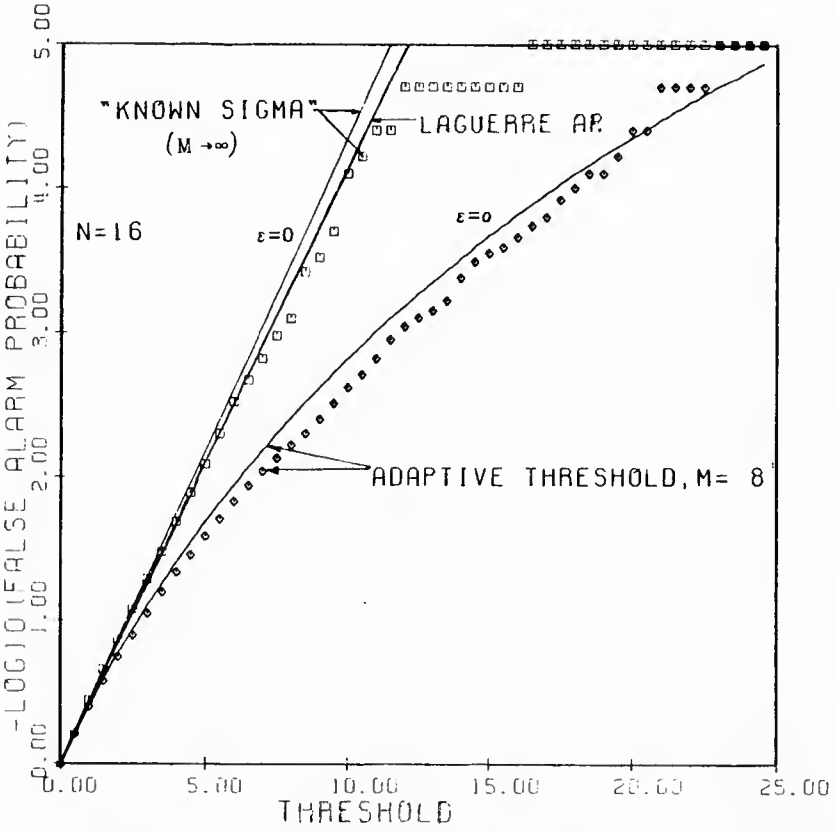


Fig. 2.2 false-alarm probability vs. threshold multiplier. Normal-Normal mixture, ε=0.01,c=3.162,k=2.025,α= .9405,50,000 monte-carlo runs

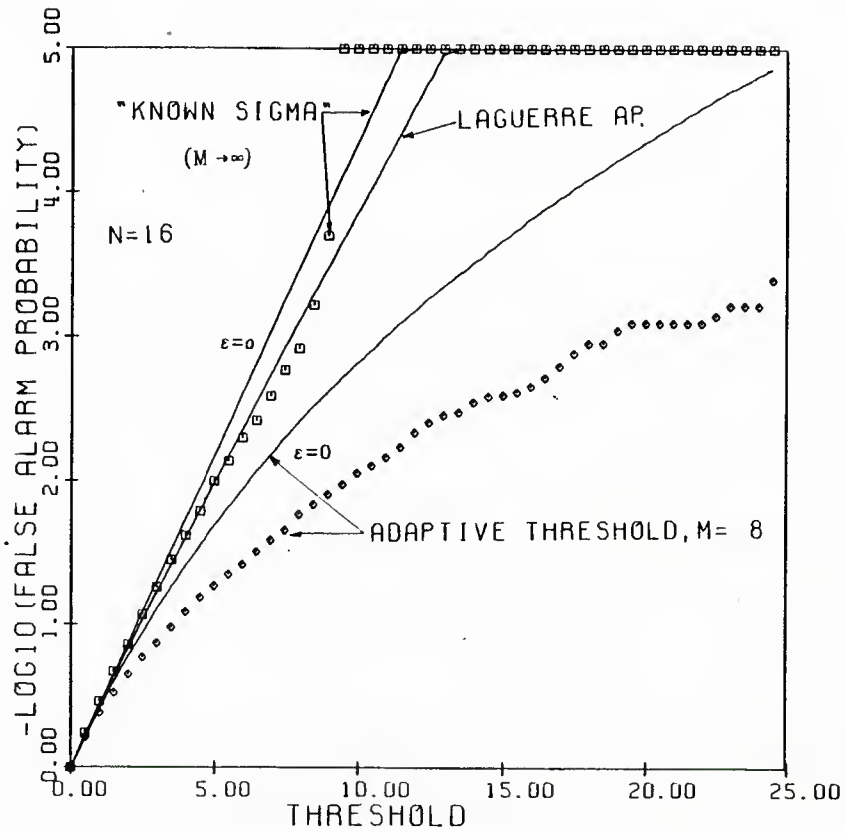


Fig. 2.3 false-alarm probability vs. threshold multiplier. Normal-Normal mixture, ε=0.1,c=3.162,k=6.06,α= .8408,10,000 monte-carlo runs

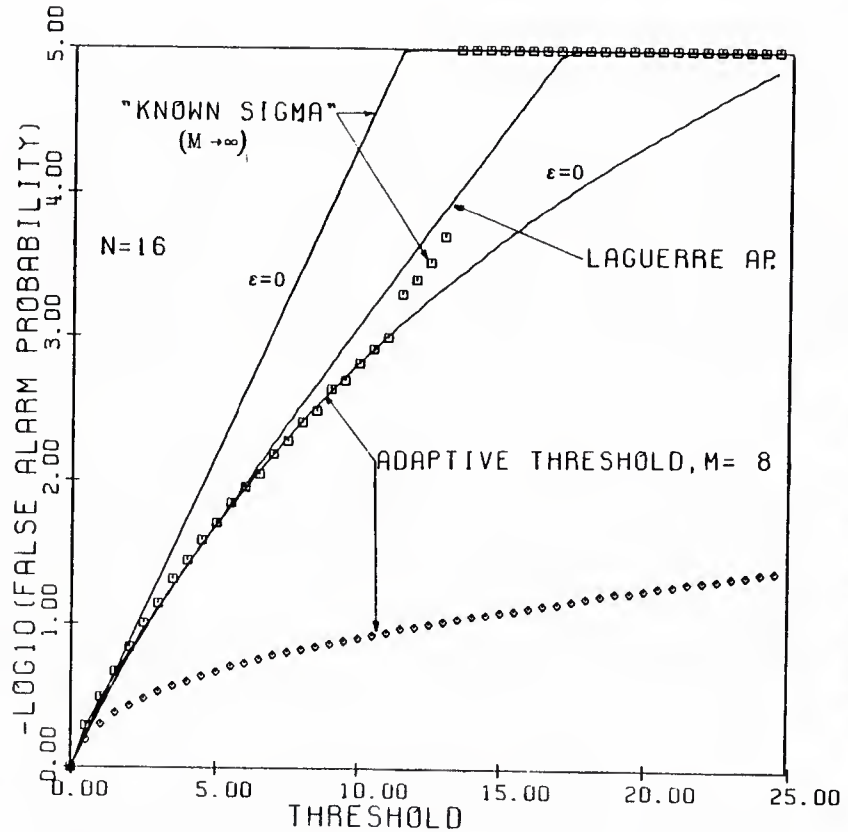


Fig. 2.4 false-alarm probability vs. threshold multiplier. Normal-Normal mixture, ε=0 1,c=10.,k=22 27,α=.5896,10,000 monte-carlo runs

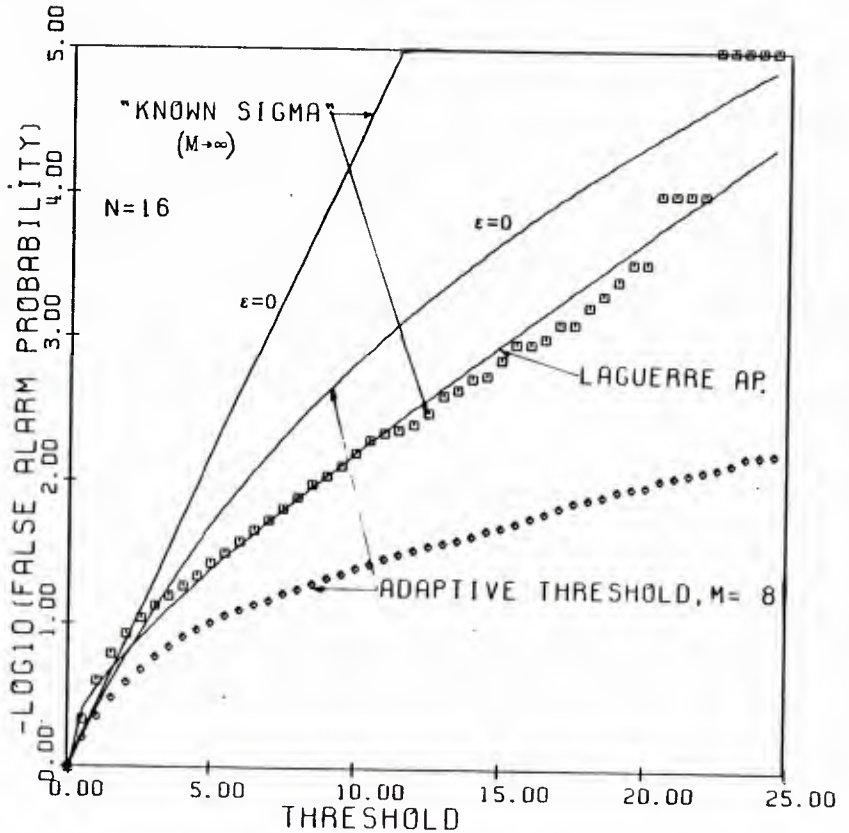


Fig. 2.5 false-alarm probability vs. threshold multiplier. Normal-Normal mixture, ε=0 01,c=10.,k=73.51,α=.3033,10,000 monte-carlo runs.

3. ROBUST TEST ON THE SAMPLE ENVELOPE

When non-contaminated hypotheses contain nuisance (vector) parameters θ_i with *known* distribution they are actually simple hypotheses, and the optimal test is a Neyman-Pearson test on the likelihood ratio of the averaged densities:

$$L(\mathbf{x}) = E_{\theta_1}[f_1(\mathbf{x} \mid \theta_1)]/E_{\theta_0}[f_0(\mathbf{x} \mid \theta_0)] \tag{3.1}$$

An attempt to incorporate averaging into the proof of optimality of Huber's test fails due to the fact that the marginal density is in general not a product of densities of the components x_i: $E_\theta[f(\mathbf{x} \mid \theta)] \neq \prod_{i=1}^{n} E_\theta[f(x_i \mid \theta)]$, unless $n = 1$. When $n = 1$, the structure of Huber's test extends to the case where there are nuisance parameters; the only correction needed is to replace $L(x)$ of Section 1.2 by $\bar{L}(x) = E_\theta[f_1(x \mid \theta)]/f_0(x)$, provided it is a monotone function of x.

Therefore, in the absence of an optimal procedure for the design of a robust test on the original observables $\{x_i, y_i\}_{i=1}^n$ of Eq. (2.7), it is proposed to weaken the requirements and to consider instead a most robust test on some *scalar* statistic. Since for $\epsilon = 0$ the sample envelope $R(\mathbf{x}, \mathbf{y}) = (\sum x_i)^2 + (\sum y_i)^2$ is a sufficient statistic for the problem, and its distribution is independent of the phase, it is natural to try to robustify a test based on R.

Thus we will consider the modified problem:

$$\mathbf{P}_0 = \{q(R) \in \mathbf{M} \mid q(R) = (1 - \epsilon')\, f_0(R) + \epsilon' c_0(R),\ \ c_0 \in \mathbf{M}\}$$
$$\mathbf{P}_1 = \{q(R) \in \mathbf{M} \mid q(R) = (1 - \epsilon')\, f_1(R) + \epsilon' c_1(R),\ \ c_1 \in \mathbf{M}\}$$

where ϵ' is the amount of contamination on the envelope R.[1]

---

[1] If one knows how to find ϵ_i of the original observables (by physical reasoning, estimation or guessing), he should be able to apply his procedure for finding ϵ'. Nevertheless, it can be verified that if the uncertainty in the measure is given by a mixture model, non-linear operations on random variables do not change ϵ and summation of n i.i.d. contaminated r.v's results in $\epsilon_n = 1 - (1 - \epsilon_i)^n$, from which we get $\epsilon' = 1 - (1 - \epsilon_i)^{2n} \approx 2n\,\epsilon_i$.

We now apply this approach to Gaussian contaminated quadrature noise components and Rayleigh distributed signal amplitude.[2] It is well known [13] that the nominal densities are given by:

$$f_1(R \mid A)=\sigma_R^{-2}\exp[-\sigma_R^{-2}(R+n^2A^2)]I_0(A\sqrt{R}/\sigma^2)\ ,\ R\geq 0 \tag{3.2}$$

$$f_1(R)=E_A[f_1(R \mid A)]=[\sigma_R^2(1+S_n)]^{-1}\exp[-R/\sigma_R^2(1+S_n)]\ ,\ R\geq 0 \tag{3.3}$$

where $\sigma^2=E_{f_0}(n_c^2)=E_{f_0}(n_s^2)$, $\sigma_R^2=2n\sigma^2, S=A_0^2/2\sigma^2$ is the input signal to noise ratio under the nominal conditions, $A_0^2=E(A^2)$ and $S_n=nS$.

The least favorable pair is given by Eqs.(1.6-1.7), with $f_1(R)$ from (3.3) and $f_0(R)$ by the same expression with $S_n = 0$, and

$$\bar{L}(R)=(1+S_n)^{-1}\exp(KR)\ ,\ R\geq 0,\quad K\triangleq S_n/\sigma_R^2(1+S_n) \tag{3.4}$$

As $\bar{L}(R)$ is monotone, an equivalent test statistic is

$$T(R)=\begin{cases}\sigma_R^2 r' & , \quad R\leq\sigma_R^2 r'\\ R & , \quad \sigma_R^2 r'<R\leq\sigma_R^2 r''\\ \sigma_R^2 r'' & , \quad R>\sigma_R^2 r''\end{cases} \tag{3.5}$$

where the limiter breakpoints are found by solving the normalization equations (C.2-C.3) of Appendix C. These equations yield

$$\exp[-r'(1+S_n)^{-1}]+(1+S_n)^{-1}\exp[r'S_n(1+S_n)^{-1}][1-\exp(-r')]=(1-\epsilon')^{-1} \tag{3.6}$$

$$1-\exp(-r'')+(1+S_n)\exp[-r''\ S_n(1+S_n)^{-1}]\exp[-r''(1+S_n)^{-1}]=(1-\epsilon')^{-1} \tag{3.7}$$

It is shown in Appendix C that the resulting maximin test can take three different forms, depending on ϵ and the desired α_0. Roughly, when ϵ/α_0 is large, it is a randomized test where H_1 is decided with probability $c<1$ if $R>t$. For intermediate values of ϵ/α_0, it is a deterministic threshold test on R, and for small values, H_0 is decided with probability c, if $R<t$. Equation (3.7) is explicitly solved by

[2] This corresponds to a Swerling I target model [17], which is a very good description of the radar cross section (RCS) fluctuations for microwave frequencies, where the target size is much larger than the wavelength so that the quadrature components are due to summation of many independent reflectors and the central limit theorem holds - [13].

$$r'' = \ln[S_n(1-\epsilon')/\epsilon'] \tag{3.8}$$

while solutions for the first one can be found numerically from the set of solutions $\{a < 1\}$ of

$$S_n a + a^{-S_n} = (1+S_n)(1-\epsilon')^{-1} \quad , \quad a = \exp[-r'(1+S_n)^{-1}] \tag{3.9}$$

Table 3.1 shows the values of the "sufficiently small" ϵ_c' for which $\mathbf{P}_0$ and $\mathbf{P}_1$ are disjoint and the maximin test exists ($L'(\epsilon') \leq L''(\epsilon')$, $\forall \epsilon' \leq \epsilon_c'$). These are given as a function of the integrated signal to noise ratio S_n, the natural measure of distance between the hypotheses for this problem.

S_n	10^{-2}	10^{-1}	1	10	10^2	10^3	∞
ϵ_c'	$3.5 \cdot 10^{-3}$	$4 \cdot 10^{-2}$	.2	.42	.49	.498	.5

Table 3.1 Values of critical contamination ϵ_c' vs. integrated nominal signal-to-noise ratio $S_n = nE(A^2)/2\sigma^2$.

A suitable measure of performance of a robust test is how far is the lower bound on the power $\beta(d^*, q^*)$, from the power of the Neyman-Pearson test for $\epsilon = 0$, β_0. For our problem $\beta_0 = \alpha_0^{1/(1+S_n)}$, and using (3.7) with (C.11) and (C.15) of Appendix C,

$$\beta(d^*, q^*) = \begin{cases} (1-\epsilon')^{S_n/(1+S_n)} (\alpha_0 - \epsilon')^{1/(1+S_n)} & , \quad \epsilon' \leq \dfrac{S_n \alpha_0}{1+S_n} \\ \dfrac{\alpha_0}{1+S_n} \left[S_n \dfrac{1-\epsilon'}{\epsilon'} \right]^{S_n/(1+S_n)} & , \quad \epsilon' > \dfrac{S_n \alpha_0}{1+S_n} \end{cases} \tag{3.10}$$

The different expressions correspond to cases b) and c) in the appendix. The last one is valid for those desired values of α_0 when the limiter is "effective" and one must resort to a randomized test, which in turn causes faster decrease of the power curve. Case a) of the appendix is never applied in this Rayleigh signal example, as it was found numerically that it corresponds to false-alarm probabilities higher than 0.5.

Figures 3.1 - 3.2 depict $\beta(d^*, q^*)$ versus ϵ' for different values of ϵ' and S_n. All the graphs are clearly characterized by a sharp "knee" at $\epsilon' \approx \alpha_0 S_n/(1+S_n) \approx \alpha_0$, such that

when ϵ' is larger than this critical value the power of the test deteriorates rapidly. Notice that even an increase of orders of magnitude in the effective signal to noise ratio S_n does not help to alleviate the problem. The figures also show the false alarm probability of the optimal detector for the *uncontaminated case*, when it actually operates in a worst-case contamination under H_0: $\alpha(N.P.,w.c.)=\alpha_0(1-\epsilon')+\epsilon'$. (Under H_1 the worst case power is $(1-\epsilon')\beta_0$ which is insignificantly lower than β_0 for the small ϵ's considered).

Thus we conclude that the proposed test can protect against contaminations which would have roughly doubled the false alarm probability of the Neyman-Pearson test ($\epsilon' \lesssim 2\alpha_0$). Beyond that, this protection is achieved at an intolerable price of decrease of the power. In radar applications, α_0 is very small and its variation is significant only for an order of magnitude changes. Thus the test proposed in this section is not very useful in many practical situations.

A somewhat similar robust detector for a Rayleigh signal in nearly Gaussian noise was recently studied by Shin and Kassam [11]. Motivated by insight gained from the structure of various robust detectors, they suggested inserting a limiter-squarer nonlinearity at the outputs of the in-phase $I=\sum x_i$ and quadrature $Q=\sum y_i$ channels, summing the outputs and comparing to a threshold. The amount of limiting was optimized numerically and the performance was analyzed. Comparing Fig. 6 of [11] with Figs. (3.1)-(3.2) here, it is observed that the scheme of [11] is somewhat better. However, the numerical analysis of [11] was carried out *only* for the range $\epsilon \leq \alpha_0$, and it is not known whether the performance degrades further when ϵ is orders of magnitude higher than α_0.[3] At the least, this comparison demonstrates that it is advisable to "push back the limiting" (or more precisely, an adequate censoring of the outliers as will be clear from Chapter 4) as far as possible from the threshold comparison point to the original observables.

[3] Private communication with the authors.

Figure 3.1 a) False alarm probability that would have resulted when *worst-case* contamination is applied to the UMP test for purely Gaussian noise, vs. ϵ'. b) Maximin bound on the detection probability vs. ϵ'. $\alpha_0 = 10^{-4}$

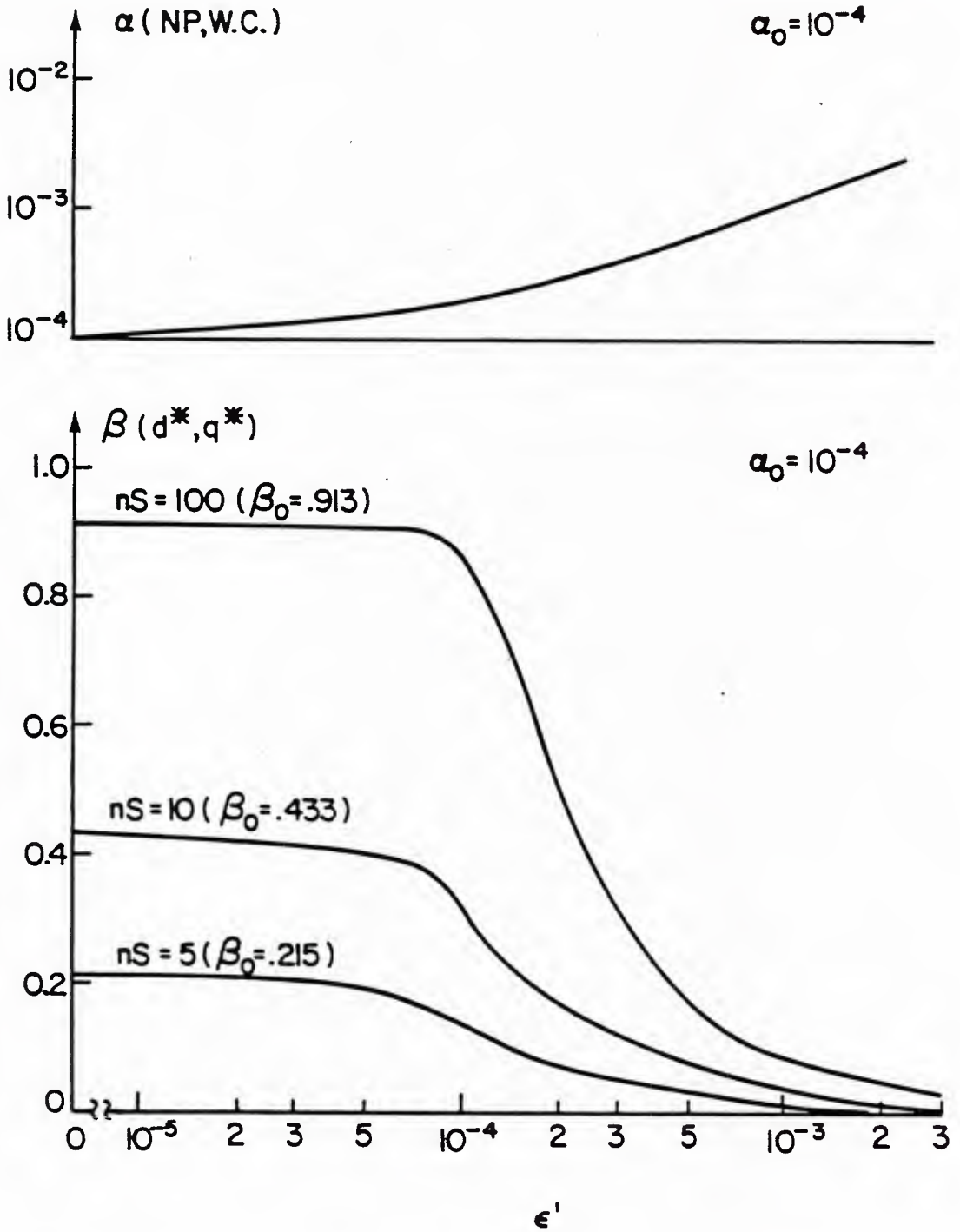
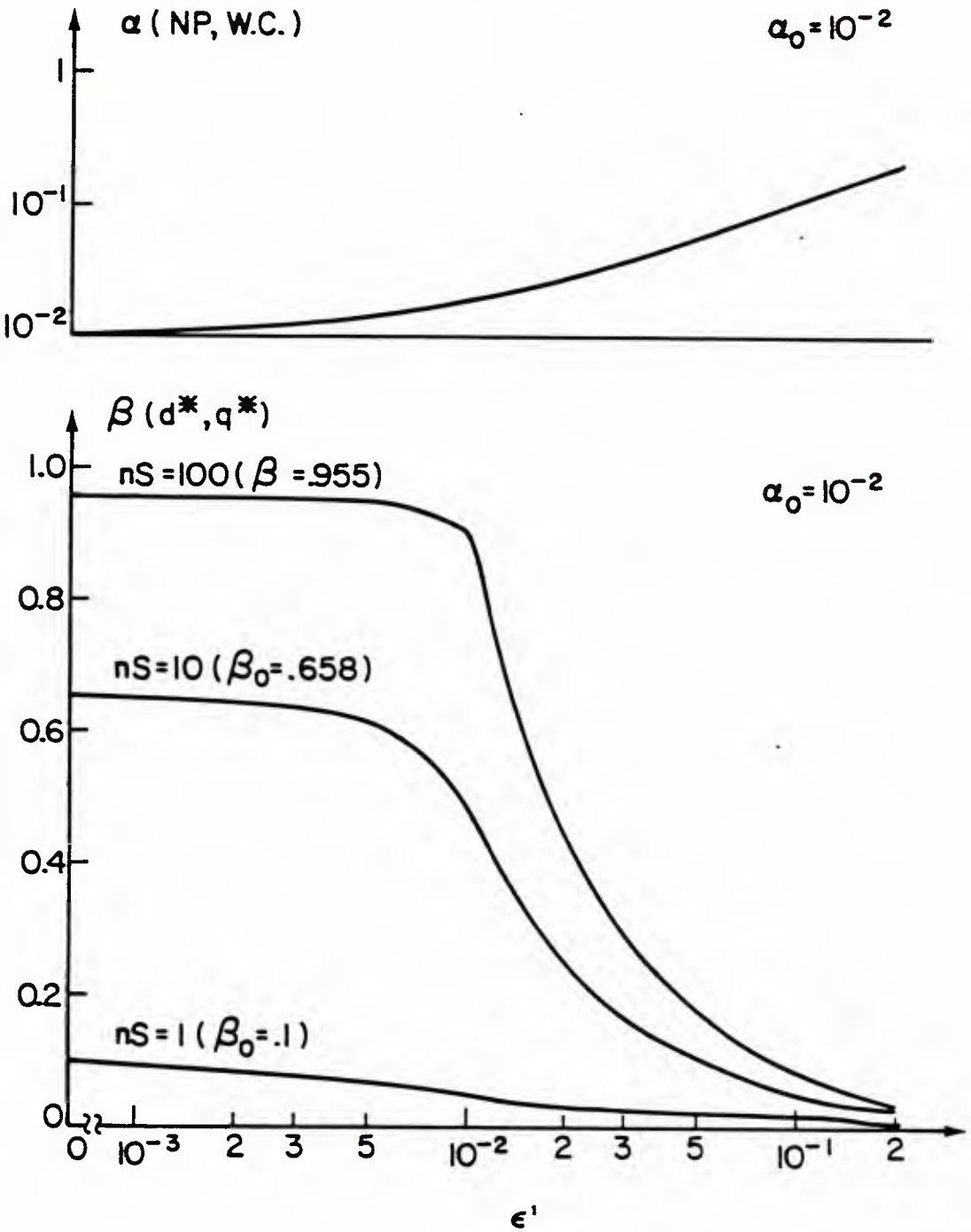


Figure 3.2 a) False alarm probability that would have resulted when *worst-case* contamination is applied to the UMP test for purely Gaussian noise, vs. ϵ'. b) Maximin bound on the detection probability vs. ϵ'. $\alpha_0 = 10^{-2}$



4. ASYMPTOTICALLY ROBUST TEST FOR KNOWN NOMINAL SCALE

4.1 Introduction

In this chapter a test will be derived which is asymptotically ($n \rightarrow \infty$) most robust in a maximin sense. The scale of the nominal distribution is assumed known and reference samples are not utilized. In the following chapter, we will show by means of simulation that this test maintains its performance even for small sample sizes.

Utilizing some of Huber's [1] results on robust estimation of a location parameter, El-Sawy and VandeLinde [6], Rieder [73], Kassam et al. [10] and others derived asymptotically robust tests for the problem of detecting a completely known signal in additive noise of uncertain distribution. A similar approach is possible (and turns out to be optimal) for the problem formulated in Section 2.1. We observe the similarity between the narrowband slowly fading coherent signal and the lowpass deterministic signal: for a *given* received sample $\{\mathbf{x}, \mathbf{y}\}$, $A\cos\phi$ and $A\sin\phi$ are essentially unknown location parameters for the quadrature signal components. Moreover, in purely Gaussian noise the unknown phase (which is a nuisance parameter) is "averaged out" in the UMP test statistic $R = I^2 + Q^2$, where I and Q are the sample-means of the in-phase and quadrature components of the narrowband observations. I and Q are the Maximum-Likelihood estimates of the component locations and thus R is a good estimate of the amplitude (squared) which is the "true" parameter of the problem, i.e., the one that distinguishes between H_0 and H_1.

Building further on insight gained from the UMP statistic R, we observe that in the purely Gaussian case the detection probability $\beta = E_A E_\phi Prob\{R > t \mid A, \phi\}$ is analytically tractable as a result of the unbiasedness and Gaussianity of I and Q under both hypotheses. Since in order to exhibit a maximin property for a composite testing problem, it is necessary to obtain analytical expressions for the error probabilities functionals, it appears natural to preserve the structure of the R test, but to replace the

sample-mean estimates with robust estimates of the locations $\{A\cos\phi,\ A\sin\phi\}$, which will have the following properties. The estimates should be:

i) Shift invariant (or unbiased) for all n.

ii) Efficient and consistent.

iii) Conditioned on A and ϕ, asymptotically Gaussian.

iv) Possess the minimax property on the estimation variance (which is translated into the effective SNR when iii is valid).

v) Maintain the above properties for medium and even small n.

Huber's M-estimator [1], and the α-trimmed mean estimator (when the nominal p.d.f. is Gaussian) have all these properties. This facilitates the proof of maximin optimality of the proposed test when the amplitude is non-random. For random amplitudes which are of more interest in practical applications, the expectation of the detection probability over the amplitude must be taken. It turns out that for the most frequently used target model in radar applications (Rayleigh amplitude or Swerling case I), the maximin property of the proposed test is only "almost" preserved (recall Fig. 1.2), but it is maintained exactly for a higher order chi-squared amplitude model. (Rayleigh is first order while constant amplitude is the limit of the chi-square family when the number of degrees of freedom tends to infinity).

For the sake of continuity and convenience in the exposition, we will begin with the non-random amplitude case, and then move to the more realistic random amplitude models. Throughout this chapter the amount of contamination ϵ in the Huber-Tukey model, as well as the variance of the nominal p.d.f., are assumed to be known. These strong assumptions which are inherent in previous work on robust detection [4]-[12], but are rarely satisfied in practice, will be relaxed in Chapter 5.

4.2 Non-random Signal Amplitude

Before presenting the test, we need some definitions. Define a class of functions Ψ on R^1 such that $dL(x)/dx = l(x) \in \Psi$ if:

1) L is convex, symmetric about the origin and increasing for positive argument.

2) l is continuous.

3) for all $q \in \mathbf{P}_0$, $0 < E_q[l^2(x)] < \infty$. (4.1)

4) for all $q \in \mathbf{P}_0$, $\frac{\partial}{\partial\theta} E_q[l(x-\theta)]$ exists and is nonzero at $\theta = 0$.

Define the estimates by the following implicit equations

$$\hat{A}_x(l) = Arg\{\sum_{i=1}^{n} s_i\, l(x_i - s_i \hat{A}_x) = 0\} \quad , \quad \hat{A}_y(l) = Arg\{\sum_{i=1}^{n} s_i\, l(y_i - s_i \hat{A}_y) = 0\} \tag{4.2}$$

($\hat{A}_x$ and $\hat{A}_y$ are the robust M-type estimates [1] for the location parameters $A\cos\phi$ and $A\sin\phi$ of the observables in Eq.(2.7b), for a given realization of the random variables A and ϕ.) Define:

$$T_n(l_0) = [\hat{A}_x(l_0)]^2 + [\hat{A}_y(l_0)]^2 \tag{4.3}$$

where

$$l_0(x) = -d[\log q_0^*(x)]/dx \tag{4.4}$$

and $q_0^* \in \mathbf{P}_0$ minimizes Fisher's information:

$$I(q_0^*) \le I(q) = \int [\frac{d}{dx} \log q(x)]^2\, q(x)dx\,, \quad \forall\, q \in \mathbf{P}_0 \tag{4.5}$$

Consider a threshold test on $T_n(l_0)$

$$d_n^*(\mathbf{x}, \mathbf{y}) = \begin{cases} H_1 \ , & T_n(l_0) \ge t_n(l_0) = t(l_0)/n \\ H_0 \ , & \text{otherwise} \end{cases} \tag{4.6}$$

and let $T_n(l)$ be any other test based on $\hat{A}_x(l)$ and $\hat{A}_y(l)$ with $l \in \Psi$ but $L \neq -log\ q_0^*$. The following proposition states the asymptotic maximin properties of the test d_n^*, which we will subsequently refer to as the SSQME (Sum of SQuared M-type Estimates) test.

Proposition 4.1 Consider the detection problem defined on the observables of Eq. (2.7) with the uncertainty in the independent quadrature noise samples obeying an ϵ– mixture model $\mathbf{P}_0$ with symmetrical p.d.f. When A is nonrandom and ϕ is uniformly distributed on $[0, 2\pi]$, among the class of all tests $T_n(l)$ based on $\hat{A}_x(l)$ and $\hat{A}_y(l)$ of Eq. (4.2) with $l \in \Psi$, there exists a least-favorable p.d.f. $q_0^* \in \mathbf{P}_0$ and a constant $0 \leq \beta_c \leq 1$ such that the SSQME test defined in (4.6) with $n \rightarrow \infty$ achieves:

$$\sup_{l \in \Psi} \beta(l, q_0^*) = \beta(l_0, q_0^*) = \inf_{q \in \mathbf{P}_0} \beta(l_0, q) \tag{4.7}$$

whenever $\beta(l_0, q_0^*) \geq \beta_c$ and subject to

$$\sup_{\substack{l \in \Psi \\ q \in \mathbf{P}_0}} \alpha(l, q) = \alpha(l_0, q_0^*) = \alpha_0 \tag{4.8}$$

Moreover, when the nominal p.d.f. in $\mathbf{P}_0$ is Gaussian, the nonlinearity l_0 is a symmetric soft limiter whose breakpoints depend on the noise variance but not on the signal amplitude:

$$l_0(x/\sigma^2, -K/\sigma, K/\sigma) = \begin{cases} K(\epsilon)/\sigma & , \quad x \geq K(\epsilon)\sigma \\ x/\sigma^2 & , \quad -K(\epsilon)\sigma \leq x < K(\epsilon)\sigma \\ -K(\epsilon)/\sigma & , \quad x < -K(\epsilon)\sigma \end{cases} \tag{4.9}$$

with $K(\epsilon)$ and ϵ related by

$$f_0(K(\epsilon))/K(\epsilon) - \Phi(-K(\epsilon)) = \epsilon/2(1-\epsilon) \tag{4.10}$$

and $f_0(x) = d\Phi(x)/dx$ is the standard normal p.d.f.

Comments

1) Note that the theorem shows the saddle-point pair of p.d.f. and test only among a certain class $T(l)$. We would like the left relation in (4.7) to be valid for *any* hypotheses test on the original observables $\{\mathbf{x}, \mathbf{y}\}$. Unfortunately, this could not be proved, unlike the case of a deterministic signal in [6], as will be clear in the sequel. However, the class $T(l)$ is quite large. As limiting cases it contains the

sample means of $\mathbf{x}$ and $\mathbf{y}$ when $\epsilon = 0$[1], the sample medians ($\epsilon = 1$) and with weak regularity conditions which usually holds [1] $\hat{A}_{\mathbf{x}}$ and $\hat{A}_{\mathbf{y}}$ can be any translation invariant statistics, such as $\hat{A}_{\mathbf{x}} = \sum_{i=1}^{n} a_i x_{(i)}$ provided $\sum_{i=1}^{n} a_i = 1$ ($x_{(i)}$ are the rank ordered samples $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$), and similarly for $\hat{A}_{\mathbf{y}}$.

2) The theorem remains valid (but l_0 given by a different expression than (4.9)) for any other description of the uncertainty in the noise, provided the asymptotic variance of the M-estimates (4.2) computed with l_0 is the maximum achievable over the class of densities in $\mathbf{P}_0$. An example is the P family [6]. Others can be found in [1], [3].

3) In contrast to Huber's finite sample size test, this asymptotic test satisfies a more desired and natural optimality criterion. Namely, the maximin relation for the power Eq. (4.7) is specified in terms of the uncertainty in $\mathbf{P}_0$ *alone,* allowing the uncertainty in $\mathbf{P}_1$ to be induced by that in $\mathbf{P}_0$ as a shift of the densities of the quadrature samples (recall the discussion in Section 1.2).

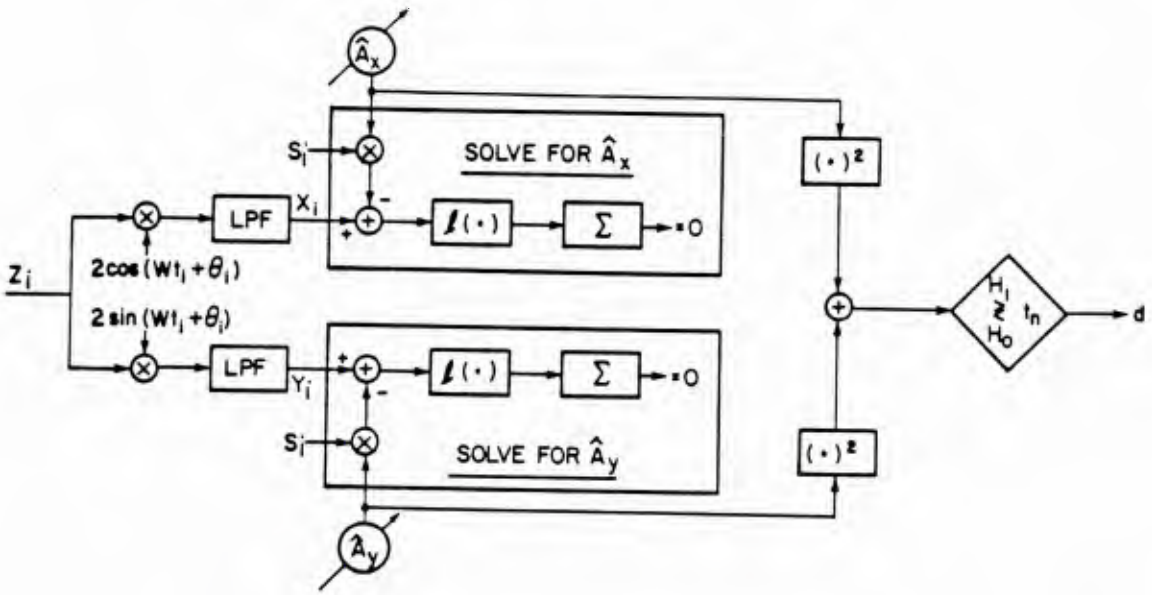
4) Extension to unknown frequency is straightforward by constructing a parallel bank of SSQME tests, the i^{th} best being matched to $w_i = i/nT$ where T is the sampling period. This will be discussed in more detail in Section 5.3. The individual test is shown pictorially in Fig. 4.1. Note that a routine for solving for a zero of a function is needed.[2] Thus, a real time implementation in a radar or sonar system calls for substantially larger number of calculations than encountered in the more common detection systems that incorporate linear filtering, FFT and at most, passing the samples through a nonlinearity.

---

[1] Thus the SSQME test is better (at least asymptotically) than the "scalar" envelope test which was derived in the preceding chapter.

[2] This is true except for the case where the nominal p.d.f. of the uncertainty family is Gaussian, where a considerable simplification is possible as will be shown later.

Figure 4.1- Block-diagram structure of SSQME test for robust detection of slow-fading narrowband signal.



Proof of proposition 4.1 We need to study the asymptotic distribution of $T(l)$, $l \in \Psi$. It can be derived using the following lemmas from [1], with a slight modification for time varying signal $\{s_i\}$ from [6].

Lemma 1 Whenever $l \in \Psi$ and q is any symmetric p.d.f., conditioned on A and ϕ, $\sqrt{n}\,(\hat{A}_x(l)-A_x)$ and $\sqrt{n}\,(\hat{A}_y(l)-A_y)$ are asymptotically distributed as normal r.v.'s with zero mean and variance $V(l, q)$

$$V(l, q) = \frac{E_q[l^2(x)]}{C\left[\frac{\partial}{\partial\theta} E_q[l(x-\theta)]\,|_{\theta=0}\right]^2}, \quad C \triangleq \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} s_i^2 \qquad (4.11)$$

where the parameters (A_x, A_y) are the true values of the locations (0 under H_0, $A_x = A\cos\phi$ and $A_y = A\sin\phi$ under H_1).

Lemma 2 Let $q \in \mathbf{P}_0$ be an ϵ-contaminated mixture p.d.f. where the nominal p.d.f. f_0 is symmetric and twice continuously differentiable, such that $-log\ f_0$ is convex on the convex support of F_0. Then, there exists $q_0^* \in \mathbf{P}_0$ which minimizes Fisher's information (4.5) over $\mathbf{P}_0$ and an $l_0 \in \mathbf{\Psi}$ given by (4.4), and the asymptotic variance of the M-estimates satisfies a saddle-point relation

$$\sup_{q \in \mathbf{P}_0} V(l_0, q) = V(l_0, q_0^*) = 1/I(q_0^*) = \inf_{l \in \mathbf{\Psi}} V(l, q_0^*) \tag{4.12}$$

Specifically, q_0^* is given by

$$q_0^*(x) = (1-\epsilon)\cdot\begin{cases} f_0(x_0)e^{K(x-x_0)} & x \leq x_0 \\ f_0(x) & x_0 < x < x_1 \\ f_0(x_1)e^{-K(x-x_1)} & x \geq x_1 \end{cases} \tag{4.13}$$

and $x_0 < x_1$ are the endpoints of the interval where $|f_0'/f_0| \leq K$, and K is related to ϵ through

$$\int_{x_0}^{x_1} f_0(x)dx + \frac{f_0(x_0) + f_0(x_1)}{K} = \frac{1}{1-\epsilon} \tag{4.14}$$

We note that the right inequality in (4.12) is a consequence of $\hat{A}_x(l_0)$ and $\hat{A}_y(l_0)$ being the maximum likelihood (ML) estimates of the corresponding location parameters, when the underlying p.d.f. is q_0^*, under *both* hypotheses. In addition, they are efficient under the assumed conditions on $\mathbf{\Psi}$, i.e., they achieve the Cramer-Rao lower bound $I^{-1}(q_0^*)$. The left inequality follows from the fact that q_0^* minimizes $I(q)$ over $\mathbf{P}_0$. Relaxation of the symmetry requirement and other noise uncertainty models are treated in [1], [3] and [6].

From Lemma 1 and the mutual independence of $\mathbf{x}$ and $\mathbf{y}$, $\hat{A}_x(l)$ and $\hat{A}_y(l)$ are independent and jointly Gaussian when ϕ is fixed. Averaging over ϕ is straightforward with transformation to polar coordinates and yields

$$f\,(\hat{A}_x(l),\,\hat{A}_y(l)\,|\,A\,) = E_\phi[f\,(\hat{A}_x(l),\,\hat{A}_y(l)\,|\,A\,,\,\phi) f\,(\phi)] \;= \cdots \tag{4.15}$$

$$= \frac{n}{2\pi V(l,\,q\,)} \exp\left[-\frac{T(l)+A^2}{2V(l,\,q\,)/n}\right] I_0\left[\frac{A\;T^{1/2}(l)}{V(l,\,q\,)/n}\right] \quad , \quad T(l) \triangleq \hat{A}_x^{\,2}(l) + \hat{A}_y^{\,2}(l)$$

where $I_0(\cdot)$ is the zeroth order modified Bessel function. Also,

$$f\,(T(l)\,|\,A\,) = \pi f\,(\hat{A}_x(l),\,\hat{A}_y(l)\,|\,A\,) \tag{4.16}$$

Now, *restricting* ourselves to the class of tests based on $\hat{A}_x(l)$ and $\hat{A}_y(l)$ as the new observables of the problem, the likelihood-ratio is $\Lambda(\hat{A}_x(l),\,\hat{A}_y(l)\,|\,A\,) = \exp(-nA^2/2V)\; I_0[nAT^{1/2}(l)/V]$ which is monotone increasing in $T(l)$ for any A, deterministic or random. Therefore, in view of the Neyman-Pearson Lemma $T(l)$ is a sufficient statistic and a threshold test based on it is UMP for H_1: $A > 0$ vs. H_0: $A = 0$.

Unlike the known lowpass signal case that was treated in [6], it can not be proved here that $T(l)$ is a sufficient statistic of the original observables $\{\mathbf{x},\mathbf{y}\}$. There, where the additive signal was non-random, use was made of theorems by Wald [23] and Chernoff [24] that a test based on the ML estimate $\hat{\theta}$ (i.e., solution to $\sum l\,(x_i - \hat{\theta}) = 0$ with $l = (-\log f\,)' \in \Psi$) is asymptotically equivalent to a NP test. In our problem, $T(l_0)$ *is not* the ML estimate of A^2 when the underlying quadrature noise p.d.f.'s are q_0^* [3]. However, its expected value is $E(T(l)) = A^2 + \frac{2}{n}\, V(l,\,q\,)$. Thus, it is asymptotically unbiased and the variance of $T(l_0)$ has a minimax property by virtue of Eq.(4.12). Indeed, it can be shown that

$$Var\,(n^{1/2}T(l)) = 4A^2V(l,\,q\,)[1 + \frac{A^2}{n}\, V(l,\,q\,)] \underset{n\,>>\,A^2}{\rightarrow} 4A^2V(l,\,q\,)$$

Hence, $T^{1/2}(l_0)$ is a good robust estimate of A, and as was discussed before, the class

[3] The MLE of A is given by solving

$$d\;\log E_\phi[\prod_{i=1}^{n}\, q_0^*(x_i - A\cos\phi)\; q_0^*(y_i\; - A\sin\phi)]/dA \;= 0$$

which is analytically intractable unless q_0^* is identically Gaussian.

$\{T(l)\}$ is quite large.

Having shown that in the restricted class of tests the optimal procedure must be a threshold test on the envelope $T(l)$, it remains to validate Eq.(4.7) subject to (4.8). Using (4.16) with the definition of Marcum's Q-function $Q(a, b) = \int_b^\infty x \exp[-\frac{1}{2}(x^2 + a^2)]\, I_0(ax)dx$, and recalling that when $l \neq l_0$ we are free to adjust the threshold $t(l)$ in (4.6) such that $\alpha(l, q_0^*) = \alpha_0$ while for the pair (l_0, q) the threshold remains $t(l_0)$, the various false-alarm and detection probabilities are as summarized in Table (4.1).

	$\alpha(l, q)$	$\beta(l, q) = Q[a = A\sqrt{n/V(l, q)},\ b = \sqrt{t(l)/V(l, q)}]$
(l, q_0^*)	$\exp[-t(l)/2V(l, q_0^*)]$	$a = A\sqrt{n/V(l, q_0^*)},\ b = b_0$
(l_0, q_0^*)	$\exp[-t(l_0)/2V(l_0, q_0^*)]$	$a = A\sqrt{n/V(l_0, q_0^*)} \triangleq a_0,\ b = \sqrt{-2\log\alpha_0} \triangleq b_0$
(l_0, q)	$\exp[-t(l_0)/2V(l_0, q)]$	$a = A\sqrt{n/V(l_0, q)} = a_0\sqrt{V(l_0, q_0^*)/V(l_0, q)}$ $b = \sqrt{t(l_0)/V(l_0, q)} = b_0\sqrt{v(l_0, q_0^*)/V(l_0, q)}$

Table 4.1- False alarm and detection probabilities for various test-p.d.f. pairs.

Note that the parameter $a^2/2$ defined by

$$\frac{a^2}{2}(l, q) = \frac{nA^2}{2V(l, q)} \tag{4.16a}$$

is the "effective integrated SNR," by analogy with the UMP detector for the purely Gaussian noise where the integrated SNR is given by the same expression with $V \to \sigma^2$.

From the first and second row of the table, by monotonicity of the Q-function with respect to the first argument (a) when the second is fixed, and since from Lemma 2 $\inf_{l \in \Psi} V(l, q_0^*) = V(l_0, q_0^*)$, we get:

$$\sup_{l \in \Psi} \beta(l, q_0^*) = \beta(l_0, q_0^*)$$

This proves the left hand side of (4.7).

From the first column of the second and third row of the table, since the exponent function is monotonic increasing when V increases, by virtue of the other inequality in Lemma 2, we get $\sup_{q \in \mathbf{P}_0} \alpha(l_0, q) = \alpha(l_0, q_0^*)$. It remains to find under what conditions $\beta(l_0, q) = Q(a_0 C, b_0 C) \geq \beta(l_0, q_0^*)$, where $C(q, q_0^*) \triangleq \sqrt{V(l_0, q_0^*)/V(l_0, q)} \geq 1$ for all $q \in \mathbf{P}_0$. Note that the Q function is monotonic in either of its arguments only when the second is fixed, but here they both change proportionally and the inequality is analytically intractable. A sufficient condition is found by solving for a_0 with a fixed b_0 those values which satisfy $d\beta(l_0, q)/dC \geq 0$, $C \geq 1$. Using known properties of the partial derivatives [15, Appendix F]:

$$\frac{\partial Q(x,y)}{\partial x} = y \exp[-\frac{x^2+y^2}{2}] I_1(xy) \quad , \quad \frac{\partial Q(x,y)}{\partial y} = -y \exp[-\frac{x^2+y^2}{2}] I_0(xy) \qquad (4.17)$$

we obtain an equivalent relation which is sufficient for the right hand side of Eq. (4.7)

$$\frac{I_1(a_0 b_0 C^2)}{I_0(a_0 b_0 C^2)} \geq \frac{b_0}{a_0} \quad , \quad \forall \quad C \geq 1 \qquad (4.18)$$

From the definition of the modified Bessel functions, I_1/I_0 is monotonic increasing from 0 to 1 when the argument increases from 0 to infinity, so it suffices to solve (4.18) for $C = 1$. From the latter argument, and since it is necessary that the slope of the probability of detection functional at $C=1$ be non-negative, this becomes a necessary condition as well. By continuity and monotonicity of both sides of Eq.(4.18), a solution exists, and upon fixing the false alarm probability ($b_0 = \sqrt{-2 \log \alpha_0}$) and by virtue of the monotonicity of the Q-function with respect to a_0, the solution of Eq. (4.18) with equality defines a critical value β_c such that when $\beta(l_0, q_0^*) = Q(a_0, b_0) \geq \beta_C$, $\inf_{q \in \mathbf{P}_0} \beta(l_0, q) = \beta(l_0, q_0^*) \geq \beta_C$. This completes the proof of proposition 4.1

Remark In the above, we have assumed independence of the in-phase and quadrature noise variates. However, the same result is valid also if they are circularly symmetric, as this property is sufficient for asymptotic independence of the quadrature M-estimates (This is a special case of Appendix D, where the distribution of a "Doppler-bank" of SSQME tests is derived). As a consequence, the SSQME test is better than the "quasi robust" narrowband test that has been proposed by Kassam [12].

Some numerical values of β_C are given in Table 4.2. It is seen that for typical false alarm levels $\beta_C \approx 0.60$.

α_0	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	$\rightarrow 0$
a_0	2.41	3.20	3.86	4.41	4.92	6.16	$\rightarrow b_0$
$\beta_C = Q[a_0, b_0]$	0.68	0.64	0.62	0.60	0.60	0.59	$Q[b_0, b_0]$

Table 4.2 - Critical values of $\beta(l_0, q_0^*)$ for which the SSQME test is maximin robust, non-random amplitude.

We note that the test in [6] for a completely known signal is asymptotically maximin robust only for $\beta_C = 0.5$ (regardless of α_0), a fact that was not directly stated there. Closer examination of the situation in [6] reveals that $\beta_C = 0.5$ is due to the symmetry of the normal p.d.f. of the test statistics around its mean under *both* H_0 and H_1. This is not the case for the SSQME test, and results in a higher value. Nevertheless, when designing a radar, sonar or communication system, the desired goal is in high quality detection. Consequently, systems are specified to achieve β=0.8–0.95, and the technical parameters which result in the required SNR are designed accordingly. Therefore, obtaining the robustness of the SSQME test only for a limited range of β is of no practical limitation.

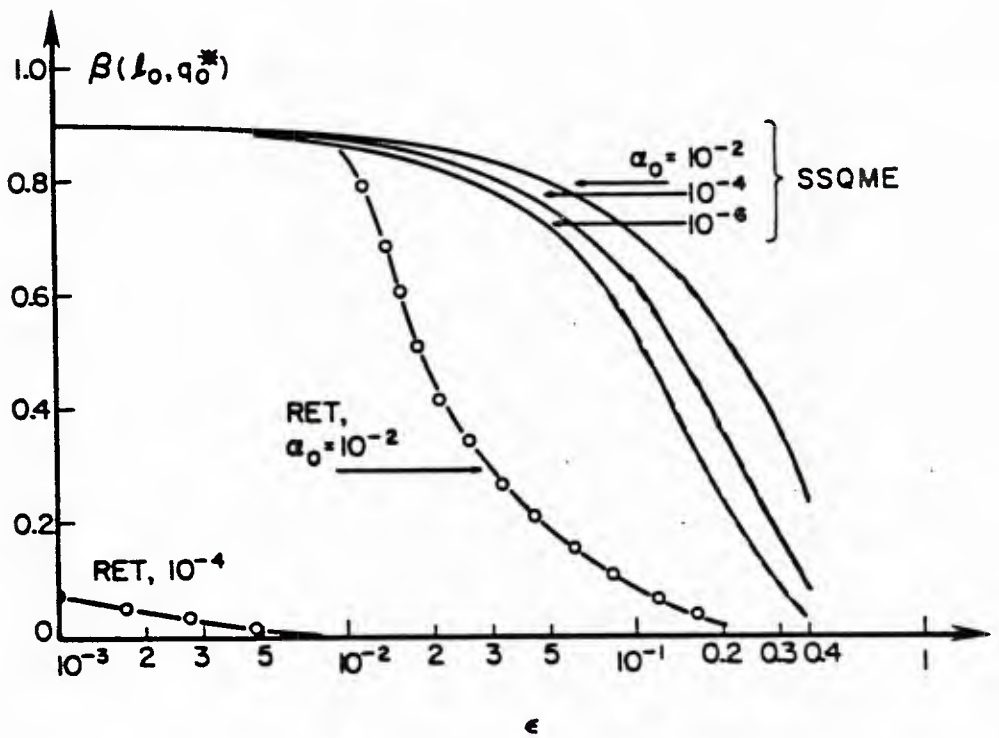
From the expression for $\beta(l_0, q_0^*)$ in Table 4.1, it is clear that the maximin bound is only slightly inferior to the detection probability in the uncontaminated case, where β is given by the same equation but with $a = A\sqrt{n}/\sigma$. This is so since for small ϵ, $rv^*(\epsilon) \triangleq V(l_0, q_0^*)/\sigma^2 \approx 1$, as can be seen in Table 4.3. (See also [3, p. 87]).

ϵ	0	.001	.002	.005	.01	.02	.05	.1	.15	.2	.25	.4
$rv^*(\epsilon)$	1	1.01	1.02	1.04	1.07	1.12	1.26	1.49	1.75	2.05	2.40	4.0

Table 4.3- Upper bound on the estimation variance (and upper bound on the loss in the effective integrated SNR), $rv^*(\epsilon) \triangleq V(l_0, q_0^*)/\sigma^2$. SSQME test for narrowband signal with non-random amplitude in nearly Gaussian noise.

The corresponding maximin bound on the detection probability is depicted in Fig. 4.2 for various values of α_0, along with the results of the robust test on the envelope of Chap. 3 (marked RET). All the curves are computed with fixed input $SNR = A^2/2\sigma^2$, such that the detection probability at the nominal Gaussian p.d.f. for all α_0's is 0.9. Very similar behavior was obtained for other SNRs. A *dramatic* improvement in performance compared to the RET is clearly observed. While the effectiveness of the RET is limited to contaminations which are smaller than roughly twice the desired false-alarm probability, (note that RET curves with $\alpha_0 < 10^{-4}$ are outside of the range of the figure) the SSQME test can protect against ϵ which is of *orders of magnitude higher* than α_0, before the detection probability decreases substantially, roughly up to $\epsilon \cong 0.1 - 0.2$.

Figure 4.2- Maximin bound on the detection probability vs. ϵ, narrowband signal with non-random amplitude and unknown phase in nearly Gaussian noise. Full lines: SSQME test. Dotted lines: RET.



4.3 Random Signal Amplitude

In radar-sonar applications, modeling targets with non-random amplitude is an oversimplification. [4] Since in nominal Gaussian noise the coherent envelope test is UMP regardless of the signal amplitude statistics, it seems natural to try to also use its robust version, the SSQME test, for random amplitudes. Therefore, we will study the robustness of the SSQME test by averaging the asymptotic power of the deterministic case with respect to the assumed amplitude p.d.f. (Obviously, the false alarm probability is bounded as before.) Two cases will be treated.

a) Swerling Case 1 Target.

This is the most common radar target, composed of many independent scatterers. The amplitude is Rayleigh, $f(A) = 2A \exp(-A^2/\overline{A^2})/\overline{A^2}$, for $A \geq 0$, $\overline{A^2} \triangleq E(A^2)$. It turns out that the performance of the SSQME test is somewhat less desirable than that for the deterministic signal case.

Proposition 4.2 Under the conditions of Proposition 4.1 and a Rayleigh distributed signal amplitude, the SSQME test with the nonlinearity l_0 of Eq.(4.4) satisfies $\beta(l_0,q) \leq \beta(l_0,q_0^*) \ \forall \ q \in \mathbf{P}_0$ subject to Eq.(4.8) - i.e., q_0^* is the least- favorable p.d.f. under H_0 but the *most* – favorable under H_1. However, when the nominal noise p.d.f. in $\mathbf{P}_0$ is Gaussian and l_0 is the soft limiter of Eq.(4.9), the right side inequality of Eq.(4.7) can be replaced by

$$\underset{q \in \mathbf{P}_0}{\beta(l_0, q)} \geq C \ \beta(l_0, q_0^*) \quad , \quad 0 \leq C \leq 1 \tag{4.19}$$

such that the constant C (which generally depends on ϵ, α_0 and $\beta(l_0, q_0^*)$) approaches unity as $\epsilon \to 0$, $\alpha_0 \to 0$ and as $\beta(l_0,q_0^*) \to 1$.

Proof Upon averaging with respect to A, the power of the SSQME test is easily found to be

[4] Except when the target is a perfect reflecting sphere.

$$\beta(l,\ q) = \exp\left[\frac{-t(l)/2V(l,\ q)}{1 + S(l,\ q)}\right] \tag{4.20}$$

where $S(l,\ q) \triangleq n\overline{A^2}/2V(l,\ q)$. Thus, with the threshold $t(l)$ set as in the first row of Table 4.1,

$$\beta(l,\ q_0^*) = \alpha_0^{1/[1 + S(l,\ q_0^*)]} \tag{4.21}$$

Clearly, $\beta(l,\ q_0^*)$ is decreasing in $V(l,\ q_0^*)$, hence $\sup_{l \in \Psi} \beta(l,\ q_0^*) = \beta(l_0,\ q_0^*)$. Unfortunately, the right side inequality of (4.7) is *never* satisfied here:

$$\beta(l_0,\ q) = \alpha_0^{1/[r + S(l_0, q_0^*)]}\ ,\ r \triangleq \frac{V(l_0,\ q)}{V(l_0,\ q_0^*)} \leq 1\ \ \forall\ q \in \mathbf{P}_0 \tag{4.22}$$

Thus $\beta(l_0,\ q_0^*)$, which is given by (4.22) with $r = 1$, is always higher. This is actually the second case that was discussed in Section 1.3.

However, the SSQME test will practically be "almost" optimal robust (in the sense that for $\forall\ q \in \mathbf{P}_0$ the performance is bounded in a narrow neighborhood: $|\beta(l_0,\ q_0^*) - \beta(l_0, q)| < \delta(\epsilon,\ \alpha_0,\ A) \rightarrow 0$, according to the discussion preceding Fig. 1.2 of Section 1.3), if the right hand side of the maximin relation (4.7) could be replaced by Eq.(4.19) such that the constant C (which generally depends on ϵ, α_0 and $\beta(l_0,\ q_0^*)$) is very close to unity. From Eq.(4.22), C is obtained for the p.d.f. for which r achieves its smallest possible value over $\mathbf{P}_0$. Some algebra yields

$$\log\ C = \left[\frac{1}{1 - r_{\min}(\epsilon)}\ \frac{\log\ \alpha_0}{\log\ \beta(l_0,\ q_0^*)} - 1\right]^{-1} \log\ \beta(l_0,\ q_0^*) \tag{4.23}$$

It is clear from this expression that $c \rightarrow 1$ monotonically as $\alpha_0 \rightarrow 0$ and as $r_{\min} \rightarrow 1$. Some basic calculus shows that this convergence is also true when $\beta(l_0, q_0^*) \rightarrow 1$. At this stage we restrict the class $\mathbf{P}_0$ to the nominal Gaussian case. Using Eqs. (4.9) and (4.11)

$$\underset{q\ =\ (1-\epsilon)g_0 + \epsilon h}{V(l_0,\ q)} = \frac{K^2 P_q\{|\ x\ | \geq K\} + \int_{-K}^{K} x^2 q(x)\,dx}{[\int_{-K}^{K} q(x)\,dx]^2} \tag{4.24}$$

where g_0 is the normal p.d.f. and $K(\epsilon)$ is given in (4.10). It is not difficult to see that (4.24) and hence also r are uniquely minimized when h is a point mass at the origin, as each of the terms in the numerator achieves its minimal value and the denominator its maximal value; using Eq.(4.24) with (4.10) yields

$$r_{\min}(\epsilon) = \inf_{q = (1-\epsilon)g_0 + \epsilon h} \frac{V(l_0, q)}{V(l_0, q_0^*)} = \frac{1 - \epsilon K^2 V(l_0, q_0^*)}{[1 + \epsilon V(l_0, q_0^*)]^2} \tag{4.25}$$

hence $r_{\min} \to 1$ as $\epsilon \to 0$.

	ϵ	.001	.01	.1	.2	.3	.5
	$K(\epsilon)$	2.63	1.945	1.14	.862	.685	.436
α_0	$r_{\min}(\epsilon)$	.9910	.9396	.6108	.349	.177	.028
10^{-2}	0.2	.995	.97	.78	.62	.52	.44
	$\beta^* =$ 0.5	.999	.994	.96	.93	.91	.89
	0.9	$1-2\cdot10^{-5}$	$1-1\cdot10^{-4}$	.999	.998	.998	.998
10^{-4}	0.2	.997	.98	.89	.81	.76	.72
	$\beta^* =$ 0.5	$1-5\cdot10^{-4}$	.997	.98	.97	.96	.95
	0.9	$1-1\cdot10^{-5}$	$1-7\cdot10^{-5}$	$1-5\cdot10^{-4}$	.999	.999	.999
10^{-6}	0.2	.998	.99	.93	.88	.84	.81
	$\beta^* =$ 0.5	$1-3\cdot10^{-4}$	.998	.986	.98	.97	.96
	0.9	$1-7\cdot10^{-6}$	$1-5\cdot10^{-5}$	$1-3\cdot10^{-4}$	$1-5\cdot10^{-4}$	$1-7\cdot10^{-4}$	.999
10^{-8}	0.2	.9987	.991	.95	.91	.88	.86
	$\beta^* =$ 0.5	$1-2\cdot10^{-4}$	.998	.99	.98	.98	.97
	0.9	$1-5\cdot10^{-6}$	$1-3\cdot10^{-5}$	$1-2\cdot10^{-4}$	$1-4\cdot10^{-4}$	$1-5\cdot10^{-4}$	$1-6\cdot10^{-4}$

Table 4.4 - The constant C such that $\beta(l_0, q) \geq C\ \beta(l_0, q_0^*)$, $\forall\ q \in \mathbf{P}_0$. SSQME test for Sw.1 target in nearly Gaussian noise.

The computation of (4.23) with (4.25) is summarized in Table 4.4. We observe that for the typical small α_0's and large β_0's desired in radar-sonar systems, C is almost indistinguishable from unity, even for very large contamination ϵ. Thus, for any $q \in \mathbf{P}_0$ and for the the range of practical interest, the asymptotic power is almost equal to the maximin bound, and the SSQME test is "almost robust" as conjectured. As in the deterministic amplitude case, the performance of the SSQME test is characterized by the identical decrease in the effective SNR $S(l_0, q_0^*)/(n\overline{A^2}/2\sigma^2)$, which is very small for small ϵ. Thus the qualitative behavior as in Fig. 4.2 is repeated here, with *dramatic* (asymptotic) improvement over the test of Chap. 3.

Further thought reveals that this "almost robustness" is a consequence of always using the same fixed threshold. If it were also possible to precisely estimate the variance of the test statistic under H_0 and to adjust the threshold accordingly: $t(q^*) = -2\ln\alpha_0 V(q_0^*) \rightarrow t(q) = -2\ln\alpha_0 \hat{V}(q) \cong -2\ln\alpha_0 V(q)$, then we would obtain $\beta(q) \cong \alpha_0^{[1+S(l_0,q)]^{-1}} \geq \beta(q^*)\ \forall q \in \mathbf{P}_0$. Since in practical applications almost always the scale (variance) of the nominal noise p.d.f. is also unknown, adaptive thresholding will be unavoidable in conjunction with the robust structure to maintain CFAR. This will be treated in more detail in Section 5.2

Comparing the known lowpass signal [6] and the non-random amplitude cases with the Rayleigh amplitude case, some thought suggests that the power fails to satisfy the desired maximin relation exactly, due to an absence of a mode in the p.d.f. of the test statistic $T(l_0)$ in the present case. To verify this, we also consider the following case which does possess a mode, and represents some physical targets.

b) Swerling Case 3 Target.

This model corresponds to one dominant reflector plus Rayleigh; the amplitude p.d.f. is $f(A) = 8A^3\exp(-2A^2/\overline{A^2})/\overline{A^2}^2$ for $A \geq 0$ [17]. In this case, the desired behavior of the deterministic amplitude case is possible.

Proposition 4.3 Proposition 4.1 is valid for a Swerling 3 target model.

Proof The power of the SSQME test is found, in a manner similar to [13], by averaging the non-random amplitude result with respect to the above amplitude p.d.f., and is given by

$$\beta(l,\, q) = \left[1 + \frac{t(l)/2V(l,\, q)}{(1 + S(l,q)/2)\,(1 + 2/S(l,\, q)}\right] \exp\left[\frac{-t(l)/2V(l,\, q)}{1 + S(l,\, q)/2}\right] \quad (4.26)$$

With the threshold set according to the known p.d.f. q_0^*,

$$\beta(l,\, q_0^*) = \left[1 - \frac{\log\,\alpha_0\, S(l,\, q_0^*)/2}{(1 + S(l,\, q_0^*)/2)^2}\right] \alpha_0^{\,1/[1+S(l,\, q_0^*)/2]}$$

Denoting $S(l,\, q_0^*)/2 = y$ and differentiating,

$$\frac{d\,\beta(l,\, q_0^*)}{dy} = \frac{-\log\alpha_0\ \alpha_0^{1/(1+y)}}{(1+y)^4}\ [2(1+y) - y\log\alpha_0\,]$$

which is positive for all $y \geq 0$, hence $\sup_{l\,\in\,\Psi} \beta(l,\, q_0^*) = \beta(l_0,\, q_0^*)$ follows by virtue of (4.12). Also, with $t(l_0) = -2V(l_0,\, q_0^*)\log\alpha_0$,

$$\beta(l_0,\, q) = \left[1 - \frac{\log\,\alpha_0\, S(l_0,\, q_0^*)/2}{(r\ + S(l_0,\, q_0^*)/2)^2}\right] \alpha_0^{\,1/[r\,+S(l_0,\, q_0^*)/2]} \quad (4.27)$$

where $r \triangleq V(l_0,\, q)/V(l_0,\, q_0^*) \leq 1$ as before. Denoting $S(l_0,q_0^*)/2 = x$ and differentiating with respect to r, yields

$$\frac{d\,\beta(l_0,\, q)}{dr} = \frac{\alpha_0^{1/(r\,+x)}\,\log\alpha_0}{(r+x)^4}[x^2 + \log\alpha_0\ x\ - r^2\,] \quad (4.27a)$$

Non-negativity of the term in square brackets for all $r \leq 1$, is sufficient for $\beta(l_0,\, q) \geq \beta(l_0,\, q_0^*)$. Clearly, from the solutions of the quadratic equation in x the positive root has to be taken, hence

$$S(l_0,\, q_0^*) \geq S_C = \max_{r\,\leq 1}\{-\log\,\alpha_0 + \sqrt{\log^2\alpha_0 + 4r^2}\,\} = -\log\alpha_0 + \sqrt{\log^2\alpha_0 + 4} \quad (4.28)$$

From (4.27a), it is clear that (4.28) is also necessary for the right hand side of the saddle-point equation. (If it was not, $\beta(l_0, q)$ would have decreased at $q = q_0^*$).

The critical values are given in Table 4.5; they are somewhat higher than those of the deterministic amplitude case, but still in the range of practical interest. Since the one-dominant-plus-Rayleigh p.d.f. corresponds to Chi-squared p.d.f. of A^2 with 2 degrees of freedom, while the non-random amplitude can be regarded as the limit of the Chi-squared family when the number of degrees of freedom N tends to infinity, it is conjectured that whenever the amplitude squared belongs to this family but $N \neq 1$ (Rayleigh), there exists a β_C such that when $\beta(l_0, q_0^*) \geq \beta_C$ the SSQME test is asymptotically maximin robust, and β_C decreases when N increases. It seems, however, that each case needs a separate derivation to prove this conjecture.

α_0	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-8}
S_C	5.36	9.70	14.1	18.7	23.2	27.8	36.9
β_C	0.78	0.75	0.743	0.740	0.738	0.738	0.737

Table 4.5 - Critical values of $\beta(l_0, q_0^*)$ for which the SSQME test is maximin robust, Swerling 3 target.

4.4 Relationship with Weak Signal LORD

For a given observation sequence, the M-estimates defined by Eq. (4.2) are fixed numbers; they are expected to be close to zero under H_0 and also under H_1 when A is small, due to the robust properties of the estimator. Thus, an approximation is obtained by expanding the summands of (4.2) in a Taylor series in powers of $\hat{A}$, and then solving (4.2) using only the two leading terms. This yields

$$\hat{A} \cong \frac{\sum_{i=1}^{n} l(x_i)}{\sum_{i=1}^{n} l'(x_i)} \cong \frac{1}{n} \sum_{i=1}^{n} l(x_i) \tag{4.29}$$

The second approximation is justified as follows: For nominal Gaussian p.d.f. with the optimal robust nonlinearity l_0, the denominator just equals the number of x_i which are in $[-K(\epsilon), K(\epsilon)]$; it is quite close to n when ϵ is small (since $K(\epsilon)$ is then larger than the variance of the samples), and only weakly changes for different realizations of $\mathbf{x}$.

All of the locally-optimal robust detectors that have been studied in [4], [5], [9], [10] and [12] are based on this test statistic, and thus they all hide an "approximate robust estimator of zero." It is worthwhile to note that the weak result of [4] and [5], namely, the LORD satisfies the maximin relation Eq. (1.10) on the slope of the power function at $A = 0$ only for $\alpha \geq \alpha_{\min}(\epsilon)$ is actually a consequence of this approximation. (It should be noted that this lower bound is *much larger than false-alarm levels of any practical interest).* Though the incremental $SNR = \lim_{A \to 0} \dfrac{E_1^2(\hat{A})}{VAR(\hat{A})}$ is identical with (4.16a), the numerator and denominator are somewhat different from those stated in Lemma 1

$$E_1(\hat{A}) \mid_{A \to 0} = A \; E_q(l') + 0(A^2) \; ; \quad VAR(\hat{A}) \mid_{A=0} = \frac{1}{n} E_q(l^2) \qquad (4.30)$$

Thus there are actually two related optimization problems (for α and β), which are not parametrized by the same functional $V(l,q)$ as in the SSQME test; this leads to a different condition that must be satisfied for Eq.(1.10) to hold - see [4], [5].

The same approach could be taken for our problem. The structure of Eqs. (4.3)-(4.6) is preserved, but the M-estimators are replaced with $\hat{A}$ of (4.29). The locally-optimal robustness can be verified using similar techniques as those in [4]. Under the regularity conditions stated in [4], which require more than the conditions in (4.1), the central limit theorem is satisfied and the detection probability is again Marcum's Q-function where the parameters $\{a, b\}$ are modified according to (4.30). Using the expression for $\dfrac{\partial}{\partial a} Q(a, b)$ in Eq. (4.17), along with $I_1(0) = 0$, $I_1'(0) \neq 0$, it can be shown that the right side inequality of Eq. (1.10) is satisfied, and thus the detector is LO robust, for a limited range in the false alarm probability given by

$$\alpha_0 \geq \alpha_{\min}(\epsilon) = \min_{q \in \mathbf{P}_0} [d(q)e(q)]^{-2/|e(q)-1|} \tag{4.31}$$

where

$$d(q) \triangleq \frac{E_q(l_0')}{E_{q_0^*}(l_0')} \quad , \quad d(q) \geq 1 \quad \forall\, q \in \mathbf{P}_0 \tag{4.32}$$

and

$$e(q) \triangleq \frac{E_{q_0^*}(l_0^2)}{E_q(l_0^2)} \quad , \quad e(q) \geq 1 \quad \forall\, q \in \mathbf{P}_0 \tag{4.33}$$

where $l_0 = -(ln\ q_0^*)'$ is the locally-optimal nonlinearity. The inequality in (4.32) is valid since l_0' is zero over the interval where the least-favorable p.d.f. q_0^* of Eq. (4.13) puts all of its contamination; the one in (4.33) is valid due to the last mentioned fact on q_0^* and to the additional regularity requirement that l_0 is monotonic nondecreasing. Because of this, the false alarm bound of Eq. (1.3) is obtained also here for $\forall\, \alpha \in (0, 1)$, but not by the same derivations as for the SSQME test. Utilizing these inequalities, it is easily verified that $\alpha_{\min}(\epsilon) \in (0, 1)$. However, we have not solved (4.31) to see if the actual values are small enough to be acceptable, as we believe that the SSQME test is a preferred solution; it is neither restricted by any $\alpha_{\min}(\epsilon)$ nor limited to weak signals. This conclusion has been verified by simulation results, to be presented in Section 5.4

5. ROBUST DETECTORS FOR A REALISTIC ENVIRONMENT

The derivation of maximin robustness of the SSQME test in Chap. 4 was based on several major assumptions, all of which have been required for previous studies of robust detection [4]-[12]. The following were assumed as completely known: a) the percentage of contamination ϵ, b) the nominal p.d.f. in the class $\mathbf{P}_0$; specifically, the variance was assumed known for the nearly Gaussian case, c) the frequency of the signal to be detected. Such knowledge concerning the detection environment is usually absent in any realistic application where robust procedures are needed. This is particularly true under the detection environment of radar (sonar) systems where the clutter (reverberation) processes rapidly change in time and space. This section extends the utility of the SSQME test for such situations, and ends with a finite-sample Monte-Carlo analysis.

5.1 Sensitivity to Unknown ϵ

In many previous investigationss [4]-[12] the amount of contamination ϵ was assumed known. Clearly, this might not hold for a nonstationary and nonhomogeneous detection environment. Theoretically, when any knowledge of ϵ is lacking, the mixture family is no longer convex and is not defined well enough to allow an optimum robust test. In practice though, it might be possible to bound $\epsilon \leq \epsilon_{\max}$ based on physical considerations. (If $\epsilon_{\max}=1$, the robust estimators of the SSQME test reduce to median estimators).

An obvious but apparently unstated observation is that any density of the form $f(x;\epsilon)$ of Eq. (2.3a) belongs to the mixture family $\mathbf{P}_0(\epsilon_{\max})$ *iff* $\epsilon \leq \epsilon_{\max}$:

$$(1-\epsilon)f_0+\epsilon h = (1-\epsilon_{\max})f_0+\epsilon_{\max}[(1-\frac{\epsilon}{\epsilon_{\max}})f_0+\frac{\epsilon}{\epsilon_{\max}}h] = \cdots \tag{5.1}$$

$$\cdots \triangleq (1-\epsilon_{\max})f_0+\epsilon_{\max}h' \quad , \ (f_0,h,h') \in \mathbf{M}$$

and the expression in square brackets h' is a legitimate density because it is positive $(1-\epsilon/\epsilon_{\max} \geq 0)$ and integrates to unity.

Table 5.1 shows the asymptotic variances of the M-estimators of Eq.(4.2) when the input noise density is a Gauss-Gauss mixture and c^2 is the variance of the contaminating density. They were computed from Eq.(4.11) with the soft- limiter of (4.9-4.10). In each sub table the limiter break-point k is designed according to ϵ_{max}, and the variances are given for $\epsilon \leq \epsilon_{max}$. Also indicated in the headings are the upper bounds on the estimation variance computed with the least-favorable density. It is seen (and can be shown analytically using the properties of the Gaussian c.d.f. when solving (4.11) in conjunction with (4.14)) that the variance monotonically increases with ϵ and c to the minimax bound. A similar behavior occurs with the detection probability since it is a monotone function of the effective SNR Eq.(4.16a), which is inversely proportional to the estimation variance.

The main conclusion from the table is that, even when the test is designed for large ϵ_{max}, the variance is much smaller when actually $\epsilon << \epsilon_{max}$. An even more important conclusion is drawn from Table 5.2, which shows a very pessimistic case: the test is designed with $k(\epsilon_{max} = 0.5)$, and the entries in the table are the ratio of variances between those obtained with this design, and those that could be obtained if ϵ was actually known and $k(\epsilon)$ was used- $Var(k(\epsilon_{max}),\epsilon)/Var(k(\epsilon),\epsilon)$. It is clearly seen that even for ϵ=0.001 the efficiency loss is merely about 28%. This is translated to an increase of only 1.07 dB in the SNR required to achieve the same P_d. When ϵ_{max} can be more tightly bounded, the difference is even smaller.

Similar behavior is found when the contaminating density is the longer tailed Laplace density, as can be seen by dividing the entries of Table 5.3a by the corresponding ones in 5.3b. (Here however, the variance is not monotone with ϵ for small c).

To summarize this discussion, we have: A design with ϵ_{max} preserves the maximin robustness properties for $\forall\, \epsilon \leq \epsilon_{max}$, and the loss incurred from not knowing the actual ϵ is shown numerically to be reasonably small for the Gauss-Gauss and Gauss-Laplace mixture families considered.

Table 5.1 Estimation variances with fixed limiter $k(\epsilon_{max})$ for $\epsilon \leq \epsilon_{max}$. **Gauss–Gauss** mixture.

EPSMAX=0.005 K(EPSMAX)= 2.160 VMAX(K)=1.037

F=(1-EPS)*N(0,1)+EPS*N(0,C**2)

EPS/C	1	3	10	30	100
0.0010	1.0066	1.0097	1.0116	1.0123	1.0125
0.0020	1.0066	1.0127	1.0167	1.0180	1.0184
0.0030	1.0066	1.0158	1.0218	1.0237	1.0243
0.0040	1.0066	1.0189	1.0269	1.0294	1.0303
0.0050	1.0066	1.0220	1.0320	1.0351	1.0362

EPSMAX=0.010 K(EPSMAX)= 1.945 VMAX(K)=1.065

F=(1-EPS)*N(0,1)+EPS*N(0,C**2)

EPS/C	1	3	10	30	100
0.0020	1.0121	1.0177	1.0210	1.0220	1.0224
0.0040	1.0121	1.0233	1.0300	1.0321	1.0328
0.0060	1.0121	1.0289	1.0391	1.0422	1.0432
0.0080	1.0121	1.0346	1.0482	1.0523	1.0538
0.0100	1.0121	1.0403	1.0573	1.0626	1.0644

EPSMAX=0.050 K(EPSMAX)= 1.399 VMAX(K)=1.256

F=(1-EPS)*N(0,1)+EPS*N(0,C**2)

EPS/C	1	3	10	30	100
0.0100	1.0467	1.0695	1.0807	1.0840	1.0852
0.0200	1.0467	1.0928	1.1157	1.1226	1.1250
0.0300	1.0467	1.1165	1.1518	1.1625	1.1662
0.0400	1.0467	1.1408	1.1890	1.2037	1.2088
0.0500	1.0467	1.1655	1.2274	1.2464	1.2530

Table 5.1 (continued).

Table 5.1 (continued).

EPSMAX=0.100 K(EPSMAX)= 1.140 VMAX(K)=1.490

F=(1-EPS)*N(0,1)+EPS*N(0,C**2)

EPS/C	1	3	10	30	100
0.0200	1.0812	1.1239	1.1433	1.1490	1.1509
0.0400	1.0812	1.1685	1.2092	1.2214	1.2254
0.0600	1.0812	1.2149	1.2793	1.2989	1.3053
0.0800	1.0812	1.2634	1.3539	1.3817	1.3909
0.1000	1.0812	1.3139	1.4333	1.4706	1.4829

EPSMAX=0.200 K(EPSMAX)= 0.862 VMAX(K)=2.046

F=(1-EPS)*N(0,1)+EPS*N(0,C**2)

EPS/C	1	3	10	30	100
0.0400	1.1393	1.2221	1.2574	1.2678	1.2706
0.0800	1.1393	1.3121	1.3904	1.4142	1.4208
0.1200	1.1393	1.4102	1.5411	1.5818	1.5933
0.1600	1.1393	1.5175	1.7125	1.7748	1.7928
0.2000	1.1393	1.6350	1.9083	1.9982	2.0249

EPSMAX=0.300 K(EPSMAX)= 0.685 VMAX(K)=2.822

F=(1-EPS)*N(0,1)+EPS*N(0,C**2)

EPS/C	1	3	10	30	100
0.0600	1.1916	1.3162	1.3684	1.3838	1.3870
0.1200	1.1916	1.4580	1.5802	1.6177	1.6260
0.1800	1.1916	1.6200	1.8367	1.9058	1.9220
0.2400	1.1916	1.8061	2.1508	2.2655	2.2941
0.3000	1.1916	2.0214	2.5405	2.7219	2.7698

EPSMAX=0.400 K(EPSMAX)= 0.555 VMAX(K)=3.996

F=(1-EPS)*N(0,1)+EPS*N(0,C**2)

EPS/C	1	3	10	30	100
0.0800	1.2395	1.4087	1.4794	1.5007	1.5061
0.1600	1.2395	1.6100	1.7855	1.8409	1.8557
0.2400	1.2395	1.8518	2.1838	2.2945	2.3256
0.3200	1.2395	2.1455	2.7140	2.9160	2.9761
0.4000	1.2395	2.5064	3.4395	3.7973	3.9099

EPSMAX=0.500 K(EPSMAX)= 0.436 VMAX(K)=5.928

F=(1-EPS)*N(0,1)+EPS*N(0,C**2)

EPS/C	1	3	10	30	100
0.1000	1.2916	1.5086	1.5999	1.6267	1.6294
0.2000	1.2916	1.7789	2.0195	2.0949	2.1047
0.3000	1.2916	2.1211	2.6094	2.7747	2.8021
0.4000	1.2916	2.5626	3.4734	3.8124	3.8808
0.5000	1.2916	3.1447	4.8072	5.5038	5.6726

Table 5.2 ratio of estimation variances between designs with fixed $k(\epsilon_{max}=0.5)$ and variable $k(\epsilon)$,**Gauss–Gauss** mixture.

VAR(K(EPSMAX))/VAR(K(EPS)) EPSMAX=0.500 K(E PSMAX)= 0.436

F=(1-EPS)*N(0,1)+EPS*N(0,C**2)

EPS/C	1	3	10	30	100
0.0010	1.2897	1.2868	1.2838	1.2828	1.2824
0.0050	1.2832	1.2734	1.2645	1.2617	1.2604
0.0100	1.2762	1.2605	1.2471	1.2428	1.2409
0.0500	1.2340	1.1963	1.1687	1.1600	1.1547
0.1000	1.1946	1.1482	1.1162	1.1062	1.0987
0.2000	1.1337	1.0880	1.0582	1.0484	1.0394
0.3000	1.0839	1.0493	1.0271	1.0194	1.0117
0.4000	1.0420	1.0224	1.0099	1.0040	1.0031
0.5000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 5.3 Estimation variances for **Gauss–Laplace** mixture. a) Fixed $k(\epsilon_{max})$. b)Variable $k(\epsilon)$

EPSMAX=0.500 K(EPSMAX)= 0.436 VMAX(K)=5.928

F=(1-EPS)*N(0,1)+EPS*LAPLACE(0,C**2)

EPS/C	1	3	10	30	100
0.1000	1.1863	1.4343	1.5708	1.6172	1.6379
0.2000	1.0922	1.5996	1.9403	2.0676	2.1252
0.3000	1.0076	1.7924	2.4425	2.7139	2.8392
0.4000	0.9314	2.0189	3.1478	3.6848	3.9411
0.5000	0.8626	2.2874	4.1795	5.2347	5.7638

VARIANCES WITH OPTIMAL K(EPS)

F=(1-EPS)*N(0,1)+EPS*LAPLACE(0,C**2)

EPS/C	1	3	10	30	100
0.1000	1.0438	1.2455	1.3979	1.4570	1.4799
0.2000	1.0318	1.4711	1.8222	1.9647	2.0244
0.3000	0.9914	1.7138	2.3675	2.6536	2.7837
0.4000	0.9327	1.9821	3.1101	3.6591	3.9318
0.5000	0.8626	2.2874	4.1795	5.2347	5.7638

5.2 Robust Scale Invariant Tests

The scale σ of the nominal density in the ϵ-mixture family affects the implementation and performance of the SSQME test in two ways. First, σ must be known for construction of the optimal robust M-estimates, as the break points of the nonlinearity are functions of it, Eq.(4.14); e.g., they are equal to $\pm\sigma k(\epsilon)$ for the nominal Gaussian case. Second, the detection threshold setting $t(l_0)$ must be proportional to $V(l_0, q_0^*)$ to obtain the desired α_0 level, (see Table 4.1), and V itself is proportional to σ^2. Even if the optimal M-estimators are replaced by some other robust and scale invariant estimators of the quadrature locations, the second crucial problem must still be addressed.

In the specific case of nominal Gaussian density, which is the main interest of our work, the first problem is solved together with a substantial simplification in implementation complexity by utilizing Tukey's alpha-trimmed mean estimator instead of the M-estimator. This estimate is defined by:

$$\overline{X}_\alpha = \frac{1}{n(1-2\alpha)} \left[\sum_{i=2+\lfloor n\alpha \rfloor}^{n-1-\lfloor n\alpha \rfloor} x_{(i)} + (1+\lfloor n\alpha \rfloor - n\alpha)[x_{(1+\lfloor n\alpha \rfloor} + x_{(n-\lfloor n\alpha \rfloor)}] \right] \tag{5.2}$$

where $x_{(i)}$ are the ordered samples: $x_{(1)} \leq x_{(2)} \leq \ldots\ldots x_{(n)}$, and $\lfloor n\alpha \rfloor$ is the greatest integer in $n\alpha$ for $0 \leq \alpha < 0.5$. The estimator simply deletes $\lfloor n\alpha \rfloor$ samples from each end and then takes the weighted mean of the remaining. Its advantage over Huber's M-estimator is that solving an implicit nonlinear equation is not necessary. With the above weighting it is translation invariant for all n (and unbiased whenever $E_f(x_i)$ exists). Bickel [33] showed that it is asymptotically Gaussian and its variance for symmetric f is

$$Var(\sqrt{n}\,\overline{X}_\alpha) \triangleq \sigma_\alpha^2 = \frac{2}{(1-2\alpha)^2} \left[\int_0^{x_{1-\alpha}} x^2 f(x)\,dx + \alpha x_{1-\alpha}^2 \right] \tag{5.3}$$

where $F(x_{1-\alpha}) = 1-\alpha$. The two estimators are quite similar for proper choice of α and $k(\epsilon)$; they both sum linearly the main bulk of the samples, but Huber's estimator treat large samples and outliers by limiting them, while the α-trimmed estimator censors them

completely. The main difference is that for small n the α -trimmed will always discard some of the observations, even if they all are small and come from the nominal, while the other might not. For $n \rightarrow \infty$, however, they are equivalent in quadratic mean. The variance of Huber's estimator with the soft- limiter nonlinearity $l_0(x\,;-k\,,\,k\,)$ is

$$V(l_0,f\,) = \frac{2[\int_0^k x^2 f\,(x\,)dx\, + k^2F(-k\,)]}{[1-2F(-k\,)]^2} \tag{5.4}$$

This is identical to (5.3) when one choses $\alpha(\epsilon)=F(-k\,(\epsilon))$. Hence they both have the same asymptotic distribution with fixed F. Moreover, the trimmed-mean estimator is also minimax for ϵ-mixture family with Gaussian nominal [3], hence it can be substituted in the SSQME test without change in its properties.

Two different asymptotically scale invariant extensions of the SSQME test are next discussed. The difference between them depends on the availability of a "noise-reference" channel.

a) "Sliding-Window" robust quadrature test

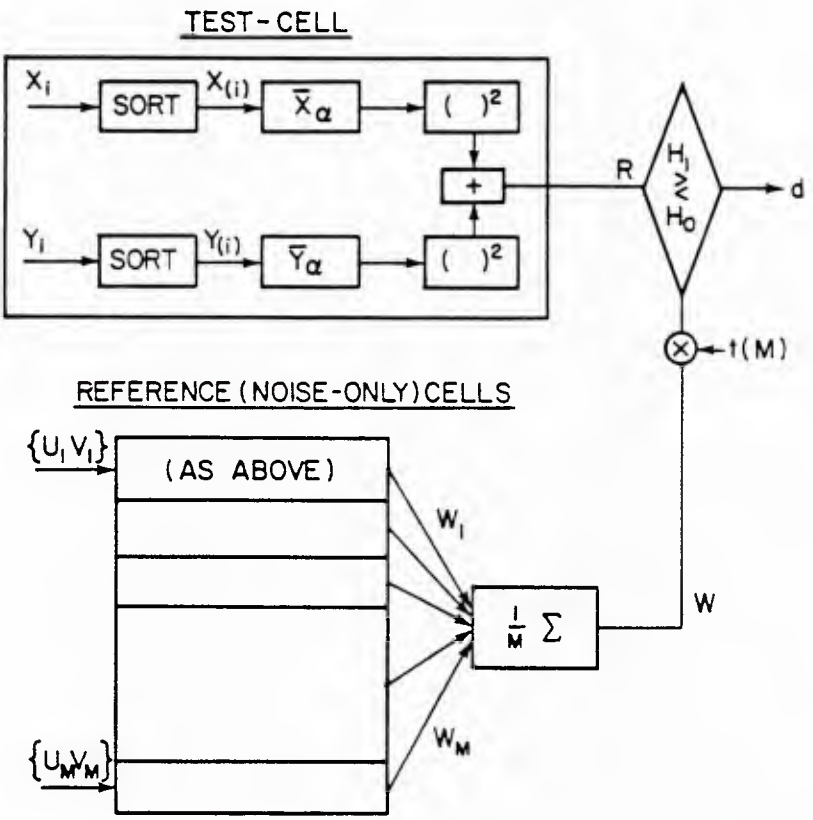
We follow the definitions and the physical setup of section 2.2. In addition to the "test-cell" samples $\{\mathbf{x},\mathbf{y}\}$, there are also available M "noise- reference" vectors of observations: $\{u_{ij}\,,v_{ij}\,\},i\,=1,...n\,,j\,=1,...,M.$ $\mathbf{u}_j \triangleq \{u_{ij}\,\}_{i=1}^{n}, j\,=1,...M$ are statistically independent vectors that are identically distributed as $\mathbf{x}$ is under H_0, and likewise for $\mathbf{v}_j$ and $\mathbf{y}$. An adaptive threshold test is constructed by using the SSQME structure in both the "test-cell" statistic $R\,(\mathbf{x},\mathbf{y})$ and in the threshold estimator $W(\mathbf{u},\mathbf{v})$ of Eqs.(2.8)-(2.11). Specifically, the test is

$$DECIDE \quad H_1 \quad if\,f \quad R_\alpha(\mathbf{x},\mathbf{y}) \quad \geq \quad t(M)W_\alpha(\mathbf{u},\mathbf{v}) \tag{5.5}$$

where $R_\alpha(\mathbf{x},\mathbf{y})$ is identical with $T_n(l_0)$ of Eq.(4.3) with the M-estimators $\hat{A}_{\mathbf{x}}$ and $\hat{A}_{\mathbf{y}}$ replaced by the α– trimmed means of the $\mathbf{x}$ and $\mathbf{y}$ samples, respectively. In a similar manner, $W_\alpha(\mathbf{u},\mathbf{v})=(1/M)\sum_{j=1}^{M} W_{j\alpha}$, where $W_{j\alpha}=\overline{U}_{j\alpha}^2+\overline{V}_{j\alpha}^2$; $\overline{U}_{j\alpha}$ and $\overline{V}_{j\alpha}$ are the α-

trimmed means of $\mathbf{u}_j$ and $\mathbf{v}_j$, respectively. The test structure is shown in Fig. 5.1.

Fig. 5.1 Sliding-Window scale invariant SSQME robust test



We can summarize the properties of this test in the following proposition:

Proposition 5.1 Consider the detection problem with the observables of Eq.(2.7), with the uncertainty in the distribution of the quadrature noise samples obeying an ϵ-mixture model $\mathbf{P}_{0\sigma}$, where the scale σ of the nominal Gaussian distribution is unknown. Let $\mathbf{D}_M$ be the class of all hypothesis tests of the form $R(T_x, T_y) \geq t_M W(T_{u_j}, T_{v_j}, j=1, \cdots M)$ where W and R are as in Eq. (5.5), with the α-trimmed estimators replaced by any other asymptotically normal, translation invariant and scale equivariant statistic T (i.e., $T(\sigma \mathbf{x}+a)=\sigma T(\mathbf{x})+a$). Let $d(\alpha) \in \mathbf{D}_M$ be the test with α-trimmed estimators, and let $\mathbf{D}_\infty$ be the class of all decision rules based on $\{T_x, T_y\}$ - i.e., without imposing the structure of Eq. (5.5) and precluding reference samples. Then:

a) The test is DF-CFAR for any $q \in \mathbf{P}_{0\sigma}$ and

$$P_{fa} = (1+t(M)/M)^{-M} \tag{5.6}$$

b) If $\alpha^*(\epsilon) = F_0^*(-k(\epsilon))$, where F_0^* is the c.d.f. of q_0^*, then q_0^* and $d^*(\alpha^*(\epsilon))$ are a saddle-point pair in $(\mathbf{P}_{0\sigma}, \mathbf{D}_M)$ for the asymptotic detection probability

$$\sup_{d \in \mathbf{D}_M} \beta(q_0^*, d) \quad = \quad \beta(q_0^*, d^*) \quad = \quad \inf_{q \in \mathbf{P}_{0\sigma}} \beta(q, d^*)$$

i.e., the test is asymptotically maximin robust. This is true for any distribution of the amplitude A, or for any amplitude level if it is deterministic. For the Rayleigh signal case, $\beta(q_0^*, d^*)$ is given by

$$\beta(q_0^*, d^*) = \left[1+\frac{\alpha^{-1/M}-1}{1+S(q_0^*)}\right]^{-M} \quad , \quad S(q_0^*) = \frac{n\overline{A^2}}{2\sigma^2} I(q_0^*(\sigma=1)) \tag{5.7}$$

c) When also $M \to \infty$, d^* is equivalent to the Neyman-Pearson test in the wider class $\mathbf{D}_\infty$ for $q_0^*(\sigma=1)$ - i.e., assuming that σ is known and only $\{T_x, T_y\}$ are available as observables.

Remarks 1) Unlike the previously studied SSQME tests, here the threshold multiplier $t(M)$ *is not* a function of the variance of the location estimators T, hence the DF-

CFAR property. 2) Unlike the cases of Chap. 4, the lower bound on the detection probability is valid for all $A>0$; in particular, this test is also maximin for the important case of a Rayleigh signal, in contrast to the non-adaptive test of Section 4.3, case a). As will be clear from the proof and from the discussion of Section 1.3, this is a result of the CFAR property. 3) The detectability loss incurred by the adaptive threshold scheme is typically small for $M>20$, cf. [18-20].

Proof a) First note that by the equivariance assumption, all tests in $\mathbf{D}_M$ are scale invariant, even for finite n. By asymptotic normality, the unknown estimation variance $V(q,T)$ appears as a common positive scale factor on both sides of Eq.(5.5), and can be divided out. Define

$$R = \frac{T_n^2(\mathbf{x})+T_n^2(\mathbf{y})}{2V(q,T)/n} \quad , \quad W=\frac{1}{2V(q,T)/n}\frac{1}{M}\sum_{j=1}^{M}[T_n^2(\mathbf{u}_j)+T_n^2(\mathbf{v}_j)]$$

Clearly, W is chi-squared distributed $f_W(w)=(1/(M-1)!)M^M w^{M-1}e^{-Mw}U(w)$, and under H_0, $f_R(r)=e^{-r}U(r)$. Since W and R are independent, $P_{fa}=P_0\{R>tW\}$ is simply computed and given by (5.6).

b) Define also $\theta=\tan^{-1}(T_n(\mathbf{y})/T_n(\mathbf{x}))$. With this transformation the density of R is computed straightforwardly as

$$f_R(r \mid A,\phi)=\int_0^{2\pi} f_{R,\Theta}(r,\theta \mid A,\phi)d\theta = \exp[-r^2-\frac{nA^2}{2V(q,T)}]I_0[A(\frac{2nr}{V(q,T)})^{1/2}] = f_R(r \mid A)$$

and in a manner similar to the derivation of Section 4.2,

$$\beta_M(q,\ T \mid A,\ w)=P_1\{R\geq tW \mid A,\ W=w\}=Q[A(\frac{n}{V(q,T)})^{1/2},(2tw)^{1/2}]$$

By monotonicity of the Q function with respect to the first argument when the second is fixed, and since A is positive, a minimax property on $V(q,T)$ is translated to a maximin property on the conditional detection probability, which in turn imply maximin on the unconditional detection probability for all $F(A)$, since

$$\beta_M(q\,,\,T)=\int_0^\infty dF(A)\int_0^\infty \beta_M(q\,,\,T\mid A\,,\,w)dF(w)$$

The derivation of (5.7) is straightforward with substitution of $F(A)$ and $F(W)$; the result is equivalent to the pure Gaussian case (compare [18-20]), if the effective SNR is changed as indicated in (5.7).

c) According to a well known convergence theorem, if $\{Y_1,\ Y_2,.....\}$ is a sequence of r.v.'s that converge in probability to $y>0$, then the distribution of the r.v. $Z_n=X/Y_n$ converges to $F_X(yz)$. Clearly, the normalized adaptive threshold estimator W converges in probability to 1 as $M\rightarrow\infty$, since it is just a sample mean. Hence, substituting $t(M)$ from (5.6),

$$\lim_{M\rightarrow\infty}\beta_M(q\,,\,T\mid A)=\lim_{M\rightarrow\infty}P_1\{\frac{R}{W}\geq t\mid A\}=\lim_{M\rightarrow\infty}Q\,[A\,(\frac{n}{V(q\,,T)})^{1/2},\,(2t(M)\cdot 1)^{1/2}]=$$

$$=\lim_{M\rightarrow\infty}Q\,[A\,(\frac{n}{V(q\,,T)})^{1/2},\,(2M(P_{fa}^{-1/M}-1))^{1/2}]=Q\,[A\,(\frac{n}{V(q\,,T)})^{1/2},\,(-2\log P_{fa})^{1/2}]$$

where the last limit follows from the dominated convergence theorem and from L'Hospital's rule. Since this expression is identical with the detection probability of the optimum asymptotic detector which operates on the transformed input observations $\{T_n(\mathbf{x}),\ T_n(\mathbf{y})\}$, see Section 4.2, the left side of the saddle-point relation is extended to the wider class $\mathbf{D}_\infty$. In particular, notice that the limit $(M\rightarrow\infty)$ of Eq.(5.7) is indeed the (exponential) detection probability of the Rayleigh case.

Remark In the proof, we did not utilize any property of the ϵ-mixture family. Therefore, the proposition is valid for any test based on asymptotically normal, shift invariant and scale equivariant estimators, whose variance is minimax in some family of distributions $\mathbf{P}$. Some examples are:

a) The set of all distributions differing at most by ϵ in Kolmogorov's distance from the Gaussian cumulative of known scale, and the corresponding M-estimator - see [3].

b) The set of all P-point distributions $\mathbf{P}=\{f : \int_{-a}^{a} f(x)dx = p\}$, and the corresponding M-estimator - see [6].

c) The set of all distributions with equal variance - this problem is interesting within the context of *optimal minimax jamming*. It turns out that the saddle-point solution is nothing but the Gaussian distribution and the quadrature test based on the sample-means. Since the saddle-point (within the sub-optimal set of SSQME tests) happens to coincide with the global UMP test at that point, it is also the minimax solution in the set of all tests - see [45]. (We also show there that the same is true even when the noise quadrature samples are not assumed a-priori identically distributed).

d) The set of all ϵ-mixture distributions with arbitrary symmetric nominal f_0 of unknown scale. Here, the α– trimmed mean is not the optimal estimator, although it is still robust in Hampel's sense as will be discussed in the following. The appropriate estimator is either from the class of L- or R-estimators[1]. As before, we take the optimal estimator for the same least favorable p.d.f. of Eq. (4.13). The location invariant and scale equivariant L-estimators are given by $L_n(\mathbf{x}) = \sum a_{ni}\, x_{(i)}$, etc., where the weights are generated from a score function m(s) according to

$$a_{ni} = \int_{(i-1)/n}^{i/n} m(s)ds \tag{5.8}$$

and $m(\cdot)$ is related to q_0^* of Eq.(4.13) through

$$m(F_0^*(x)) = \begin{cases} \dfrac{1}{I(q_0^*)} \dfrac{d^2}{dx^2}(-\log f_0(x;1)) & , \ |x| \leq x_1 \\ 0 & , \ otherwise \end{cases} \tag{5.9}$$

where

[1] The SW structure with R-estimators is an overkill. It is possible to obtain DF-CFAR performance and maximin detection probability with a quadrature test based on robustified rank statistics. This subject is studied in detail in Chapter 6.

$$f_0(x_1;1)/k - F_0(-x_1;1) = \epsilon/2(1-\epsilon)\ ,\ \ |\ f_0'\ (x_1;1)/f_0(x_1;1)\ | = k$$

Notice that also here the censoring percentage is $F_0^*(-x_1)$. The minimax relation on the estimator's variance remain valid (where $V(q_0^*) = 1/I(q_0^*)$), due to the asymptotic equivalence between L and M-estimators [Jaeckel, 46], provided the a_{ni} are chosen as above. (Stigler [47] proved later that if the nominal p.d.f is any symmetric, twice differentiable and strongly unimodal p.d.f. on R^1, the L-estimator corresponding the above score function is indeed asymptotically Gaussian with the variance assumed in [46]).

To summarize, the SW adaptive threshold extension of the SSQME test is a suitable choice for radar-sonar systems where "noise-reference" samples are conveniently available. It possesses all the desired asymptotic maximin properties; its implementation in the nominal Gaussian case is relatively easy based on α-trimmed estimators which do not require more than ordering of the data, an operation that is common to most nonparametric tests and is becoming available for real time implementation with VLSI technology [16].

b)SSQME test with a preliminary scale estimate

We assume now that {**x**,**y**} are the only available observables, and the nuisance scale estimate must be derived from the "signal plus noise " data itself [2]. We draw and build on Huber's theory [3] of simultaneous estimation of location and scale. For completeness of the presentation, we next quote and make some interpretations on Huber's main results:

i) Simultaneous M-estimates for location and scale are a pair of statistics (T_n, S_n) determined by solving equations of the form

[2] In radar/sonar applications in a nonhomogeneous environment, utilization of spatial adjacent observations (as in the SW structure) is not justified if the reference observations are not identically distributed as the main observations.

$$\sum_{i=1}^{n} \Psi\left(\frac{x_i - T_n}{S_n}\right) = 0 \tag{5.10a}$$

$$\sum_{i=1}^{n} \chi\left(\frac{x_i - T_n}{S_n}\right) = 0 \tag{5.10b}$$

These are similar to the ML estimates of θ and σ for a family of densities $\frac{1}{\sigma} f\left(\frac{x-\theta}{\sigma}\right)$, with proper selection of Ψ and χ. As most common test statistics and estimators depend on the samples only through the empirical distribution function $F_n(\mathbf{x}\,;\,x) \triangleq \frac{1}{n}\sum I_{\{x_i < x\}}$, where $I_{\{A\}}$ is the indicator function of the set A, it is convenient to express $T_n(\mathbf{x}) = T(F_n)$ and $S_n(\mathbf{x}) = S(F_n)$ in terms of the functionals defined by:

$$\int \Psi\left(\frac{x - T(F)}{S(F)}\right) F(dx) = 0 \tag{5.11a}$$

$$\int \chi\left(\frac{x - T(F)}{S(F)}\right) F(dx) = 0 \tag{5.11b}$$

Hampel's [34] influence curve (IC) is a very useful heuristic tool of robust statistics, which describes the (suitably normed) limiting influence of a single observation on the estimator. Intuitively, a qualitative robust estimator *must have a bounded IC.* If $\sqrt{n}\,[T(F_n) - T(F)]$ is asymptotically zero mean Gaussian, its variance is given by (using (5.13) below)-

$$V(F,T) = \int IC^2(x\,;F,T) F(dx) = S^2(F) \frac{\int \Psi^2\left(\frac{x}{S(F)}\right) F(dx)}{\left[\int \Psi'\left(\frac{x}{S(F)}\right) F(dx)\right]^2} \tag{5.12}$$

with a similar expression for $V(F,S)$. If (and only if) F and χ are symmetric, and Ψ is skew-symmetric, the IC's of the simultaneous estimates are uncoupled; in particular, only S(F) but neither the IC nor the asymptotic variance of S enters into the expression of IC(T). The IC's are given by:

$$IC(x\,;F,T) = \frac{\Psi\left(\frac{x}{S(F)}\right) S(F)}{\int \Psi'\left(\frac{x}{S(F)}\right) F(dx)} \tag{5.13}$$

$$IC(x;F,S) = \frac{\chi(\frac{x}{S(F)})S(F)}{\int \chi'(\frac{x}{S(F)})\frac{x}{S(F)}\ F(dx)} \tag{5.14}$$

Note that for a given distribution, T(F) and S(F) are numbers; T_n and S_n are Fisher consistent estimates if T(F) and S(F) are equal to the estimated parameters. Under the above mentioned symmetry conditions, the estimates are asymptotically uncorrelated and hence independent.

ii) The asymptotic robustness (in Hampel's sense) is clear from (5.13)-(5.14), provided that Ψ and χ are bounded functions. In Section 6.4 of [3] it is shown that under relatively mild conditions the coupled estimates are consistent, $(T_n,S_n) \rightarrow (T(F),S(F))$ in probability, and are jointly Gaussian.

iii) The IC of T for the case when the nominal scale is known is given by the same expression as (5.13) but with S(F)=1. If S(F) does not change too much when $F \in \mathbf{P}_0$ (a property which must be satisfied for any reasonably consistent and robust estimate of the scale), the desired properties of the location estimate with known scale will be roughly preserved. However, it is not possible to obtain an exact minimax, since the loss of the translation invariance symmetry in the multi-parameter case does not enable extension of the parametrization of the nominal model throughout a convex uncertainty neighborhood, cf. [3, Section 11.1]. It should also be clear that the variational techniques of [3] are not applicable for finding the density that minimizes $V(F,T)$, even when the asymptotic variance of S is considered as a nuisance factor, due to the deep and implicit nonlinear coupling of Eq.(5.12) subject to Eqs.(5.11).

iv) The simultaneous solution of Equations (5.10) is perhaps an unnecessary complicated. Simplified variants are the *one step M-estimates*. They are obtained by starting with some preliminary estimates $T_n^{(0)}$ and $S_n^{(0)}$, and then solving Eq.(5.11a) approximately by applying Newton's method *just once*. The following variant

emerged in the extensive "Princeton Monte-Carlo study " of Andrews et al. [35] as very robust, even for small sample sizes. The preliminary estimates are the median: $T_n^{(0)}(\mathbf{x}) = med\,\{x_i\}$, and the median of the absolute deviations from it (MAD): $S_n^{(0)}(\mathbf{x}) = med\,\{\mid x_i - T_n^{(0)}(\mathbf{x})\mid\}/0.6745$, where $0.6745 \cong \Phi^{-1}(3/4)$ to get consistency at the Gaussian ($\chi(x) = sign(\mid x \mid -1)$, thus $S^{(0)}(F) = F^{-1}(3/4)$ for a symmetric density from (5.11b)). The MAD is the limiting case ($\epsilon \rightarrow 1$) of the minimax robust scale estimate, and its computation is simpler. Then,

$$T_n^{(1)}(\mathbf{x}) = T_n^{(0)}(\mathbf{x}) + \frac{\frac{1}{n}\sum\Psi(\frac{x_i - T_n^{(0)}}{S_n^{(0)}})S_n^{(0)}}{\frac{1}{n}\sum\Psi'(\frac{x_i - T_n^{(0)}}{S_n^{(0)}})} \tag{5.15}$$

$T_n^{(1)}$ is asymptotically ($n \rightarrow \infty$) equivalent to the full solution of Eq.(5.10a) $T_n^{(\infty)}$, provided the previously mentioned symmetry conditions are satisfied. Moreover, with these symmetry conditions, any one-step estimator with translation invariant and odd $T_n^{(0)}$: $T_n^{(0)}(\mathbf{x}+c) = T_n^{(0)}(\mathbf{x}) + c$, $T_n^{(0)}(-\mathbf{x}) = -T_n^{(0)}(\mathbf{x})$), will have the same IC (and hence identical asymptotic properties) as the full solution of the coupled equations. Note that for the nominal Gaussian case the computation of (5.15) is of the same order of complexity as of the nonlinear detectors typical of the LORD approach (Section 1.3), provided $T_n^{(0)}$ and $S_n^{(0)}$ are available after two orderings of the data (the denominator is just the number of normalized samples in $[-k\,,k\,]$).

With this background , it seems natural to preserve the structure of the SSQME test, but with scale invariant location estimators and an adaptive threshold which is derived from the scale estimate. Thus we propose the following test:

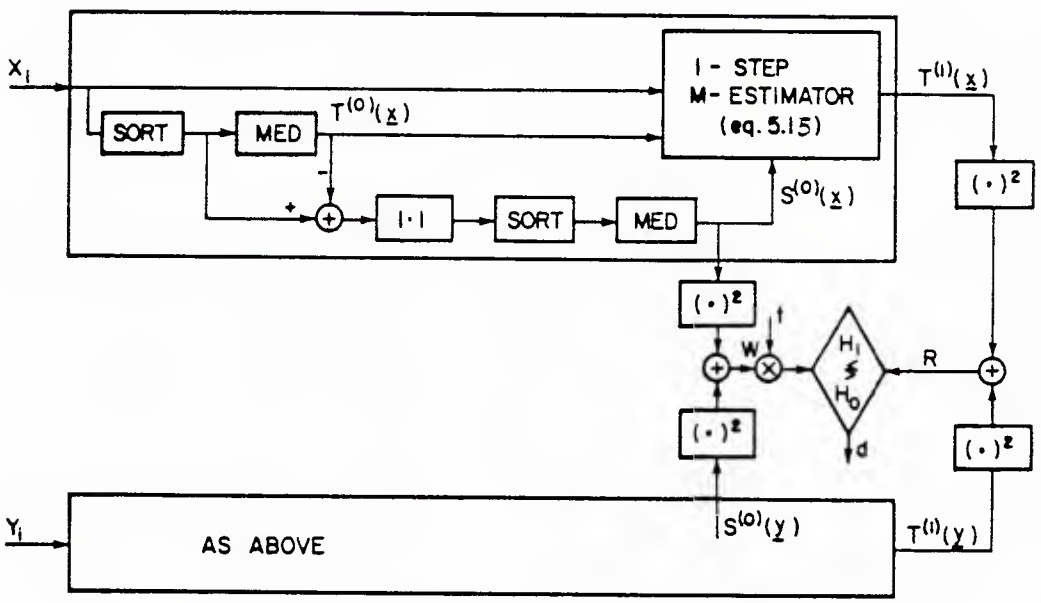
$$DECIDE \quad H_1 \quad if\,f \quad R(\mathbf{x},\mathbf{y}) \geq t_n\, W(\mathbf{x},\mathbf{y}) \tag{5.16}$$

where

$$R(\mathbf{x},\mathbf{y}) \triangleq [T_n^{(1)}(\mathbf{x})]^2 + [T_n^{(1)}(\mathbf{y})]^2 \quad , \quad W(\mathbf{x},\mathbf{y}) \triangleq [S_n^{(0)}(\mathbf{x})]^2 + [S_n^{(0)}(\mathbf{y})]^2 \tag{5.17}$$

$T_n^{(1)}$ is the one step M-estimator of Eq.(5.15) and $S_n^{(0)}$ is the MAD scale estimator. The structure of the test is shown in Fig. 5.2.

Fig. 5.2 SSQME test with MAD preliminary scale estimate.



The asymptotic properties of the test are summarized in the following. They are valid for all simultaneous and one step M-estimators provided the symmetry properties that were mentioned before hold.

Proposition 5.2 a) The test (5.16) is CFAR for any fixed density $f \in \mathbf{P}_0$ with unknown (variable) σ. b) With variable f , the changes in the asymptotic false-alarm probability are governed only by the normalized variance of the location estimates : $V'(F,T^{(1)}) \triangleq V(F,T^{(1)})/(S^{(0)}(F))^2$, but not by the variance of the scale estimate. c) The asymptotic detection probability is given by the same expression as for a fixed threshold SSQME test: $\beta(F)=\alpha(F)^{1/(1+SNR(F))}$ (for a Rayleigh signal), where

$$SNR(F) \;=\; \frac{nA_0^2}{2[S^{(0)}(F)]^2 V'(F,T^{(1)})} \tag{5.18}$$

is also not a function of the scale estimator variance.

Proof 1) Both $T_n^{(1)}$ and $S_n^{(0)}$ are scale equivariant: $S_n^{(0)}(\sigma\mathbf{x})=\sigma S_n^{(0)}(\mathbf{x})$, etc., from which a) follows immediately. Also, $S^{(0)}(F_\sigma) = \sigma S^{(0)}(F_{\sigma=1})$, hence we will omit the scale subscript on F for convenience.

2) The probability of acceptance of H_1 is given by $\int_0^\infty f_W(w)dw \int_{i_n w}^\infty f_R(r)dr$, since R and W are asymptotically independent when the symmetry conditions of i) above are met. R is one sided exponentially distributed, with

$$E_0(R) \;=\; \frac{2}{n}(S^{(0)}(F))^2 V'(F,T^{(1)}) \quad , \quad E_1(R) \;=\; E_0(R)[1+SNR] \tag{5.19}$$

by straightforward calculation. Note that the definition of SNR is justified (by analogy to the classical case) as $V'(F,T^{(1)})$ is also σ invariant by its definition, as long as the scale and location estimates are also invariant. Hence the input noise variance affects SNR only through $S^{(0)}(F_\sigma)=\sigma F^{(0)}(F_1)$. As W is the sum of the squares of two independent identically distributed Gaussian r.v.'s $N(a,v^2)$, it has a *noncentral* χ^2 distribution with 2 degrees of freedom and noncentrality parameter $\lambda=2a^2$, c.f. [13]. Explicitly,

$$f_W(w)=\frac{1}{2v}\exp(-\frac{w+2a^2}{2v})I_0(\frac{a\sqrt{2w}}{v}) \tag{5.20}$$

where I_0 is the modified Bessel function of zeroth order. Here $a=S^{(0)}(F)$ and $v=\frac{1}{n}V(F,S^{(0)})$ under H_0 and under H_1, as $S_n^{(0)}(\mathbf{x}+c)=S_n^{(0)}(\mathbf{x})$. The desired probabilities are thus computed from:

$$p_i \quad = \quad \frac{\exp(-a^2/v)}{2v}\int_0^\infty \exp[-w(\frac{1}{2v}+\frac{t_n}{E_i(R)})]I_0(\frac{a\sqrt{2w}}{v})dw \quad , \quad i=H_0, H_1 \tag{5.21}$$

The evaluation of this integral is facilitated by

$$\int_0^\infty \exp(-x)I_0(2\sqrt{yx})dx \quad = \quad \exp(y) \tag{5.22}$$

which can be derived by a change of variables from the normalization of the Rician density. The result is

$$p_i \quad = \quad \frac{\exp[-\frac{2a^2t_n}{E_i(R)}(1+2t_n v/E_i(R))^{-1}]}{1+2t_n v/E_i(R)} \quad , \quad i=H_0, H_1 \tag{5.23}$$

As $a^2/E_i(R)\approx n$ and $v/E_i(R)$ is not a function of n, in order to get a non-zero α as $n\to\infty$ we must have $t_n=t/n$. Thus when $n\to\infty$ the terms that contain $V(F,S^{(0)}(F))$ through v cancel out and $\lim_{n\to\infty} p_0 = \exp(-t/V'(F,T^{(1)})$ follows and proves b). The proof of c) is similar upon substitution of $E_1(R)$ from Eq.(5.19) into (5.23) and taking the limit as $n\to\infty$.

Comments and interpretation

1)Since the error probabilities are not functions of the variance of the scale estimator, $S_n^{(0)}$ should be chosen to achieve better and flatter consistency over $\mathbf{P}_o$, and not according to its variability. 2) The false alarm probability with the preliminary estimate is parametrized by $V'(F,T^{(1)})$, which is roughly equal to the variance under the known scale case, if $S^{(0)}(F)\cong 1$ and it does not change too much over $\mathbf{P}_0$. Hence we have a simi-

lar situation as in Chapter 4, and the upper bound on the false alarm probability is approximately satisfied if the least favorable density q_0^* is chosen. While it does not seem possible to give an analytic characterization of the density that maximizes $V'(F,T^{(1)})$ over the ϵ–mixture family $\mathbf{P}_0$ as was discussed in iii) above, if the noise density is restricted to a narrower family (e.g., Gauss-Gauss or Gauss-Laplace ϵ– mixtures), the supremum of the variance can be found (at least numerically) and adjusted accordingly to obtain an upper bound on the α over the restricted family. 3)With that restricted maximization, an approximate upper bound on the detection probability will also be obtained as follows. Denote by $\mathbf{P}_{0r}$ the restricted noise family. Since $S^{(0)}(F,D)=F^{-1}(3/4)/D$ (for the MAD estimator) changes over $\mathbf{P}_{0r}$, its arbitrary normalization D can be used as an optimization parameter. Let $V^{**}(D)=\underset{\mathbf{P}_{0r}}{Max}V'(F,T^{(1)},D)$ and $\alpha_0=\exp(-t/V^{**}(D))\geq\alpha(F,D)$. Under the same approximation of Section 4.3 for sufficiently high desired probability of detection ($\beta\gtrsim 0.5$), with substitution of SNR from Eq. (5.18):

$$\beta(F,D) \;=\; \alpha(F,D)^{1/(1+SNR(F))} \;\cong\; \alpha_0^{2[S^{(0)}(F,D)]^2V^{**}(D)/nA_0^2} \tag{5.24}$$

Hence the additional detectability loss from scale estimation relative to the worst case where σ is known is characterized by the effective SNR loss

$$L(F,D) \;=\; \frac{V^{**}(D)[S^{(0)}(F,D)]^2}{V(l_0,q^*)} \tag{5.25}$$

and D^{**} can be chosen to minimize the maximum loss over $\mathbf{P}_{0r}$: $D^{**}=\underset{D}{ArgMin}\{\underset{f\in\mathbf{P}_{0r}}{Max}L(F,D)\}$. Hence, α (β) is bounded from above (below) over the family, and the additional loss due to scale estimation is low if $L(F^{**},D^{**})\approx 1$.

Figures 5.3a-5.3f show the results of such a numerical optimization over the Gauss-Gauss mixture family, vs. the r.m.s. power of the contamination c. In each figure the dashed lines are of $V'(F,T^{(1)})/V^{**}$, and the continuous lines are of $L(F,D)$. One set of graphs in each figure corresponds to the optimization with respect to D, while the other

was obtained where D was taken to get consistency at the least-favorable density of the known σ

$$D^* = \begin{cases} k - \frac{1}{k} \ln \left[\frac{k}{4(1-\epsilon) f_0(k)} \right] & , \quad \frac{f_0(k)}{k} \geq \frac{1}{4(1-\epsilon)} \\ \Phi^{-1} \left[\frac{1 + 2(1-\epsilon)}{4(1-\epsilon)} \right] & , \quad \frac{f_0(k)}{k} < \frac{1}{4(1-\epsilon)} \end{cases} \tag{5.26}$$

The figures demonstrate that there is no *additional* loss (above that incurred from $V(l_0, q_0^*)/\sigma^2$) from the scale estimation for $\epsilon \leq 0.2$, that the loss is negligible (about 4%) for $\epsilon = 0.3$, and it is quite small (about 26%,corresponding to 1 dB) for $\epsilon = 0.4$. Even for the huge uncertainty $\epsilon = 0.5$ the loss is reasonable (50%) if $c \leq 20$. For $\epsilon > 0.5$, it seems that $L(F, D^{**})$ is unbounded. This agrees with the "breakdown" point of the joint estimation, as evaluated by Huber [3, pp. 141-146]. Also, the loss difference between the optimal D^{**} and the normalization D^* is quite small. As $D^{**} > 0.6745$, the efficiency at the nominal Gaussian is somewhat sacrificed to obtain better performance over $\mathbf{P}_{0r}$ (compare with Table 5.1). Similar conclusions can be drawn from Figures 5.4a-5.4d which are valid for the Gauss-Laplace mixture family. When the nominal p.d.f. is Gaussian, the test of Eq.(5.17) can be simplified by utilizing α-trimmed estimators in place of the one step location estimator. In that case, it turns out that the agreement with the performance of the known scale is even closer, see [48] for the numerical data.

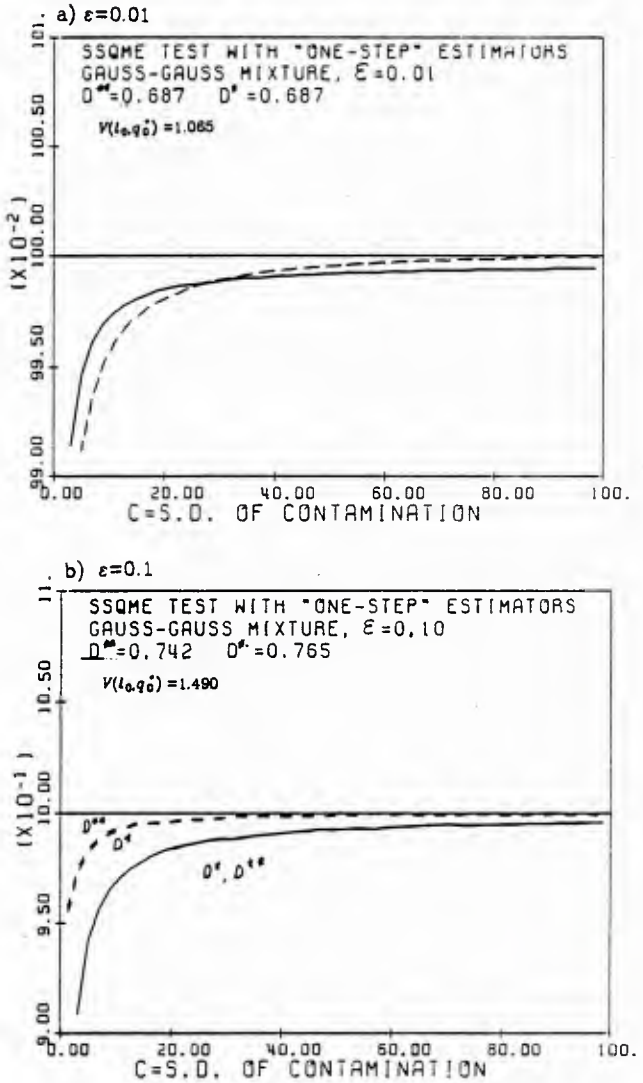
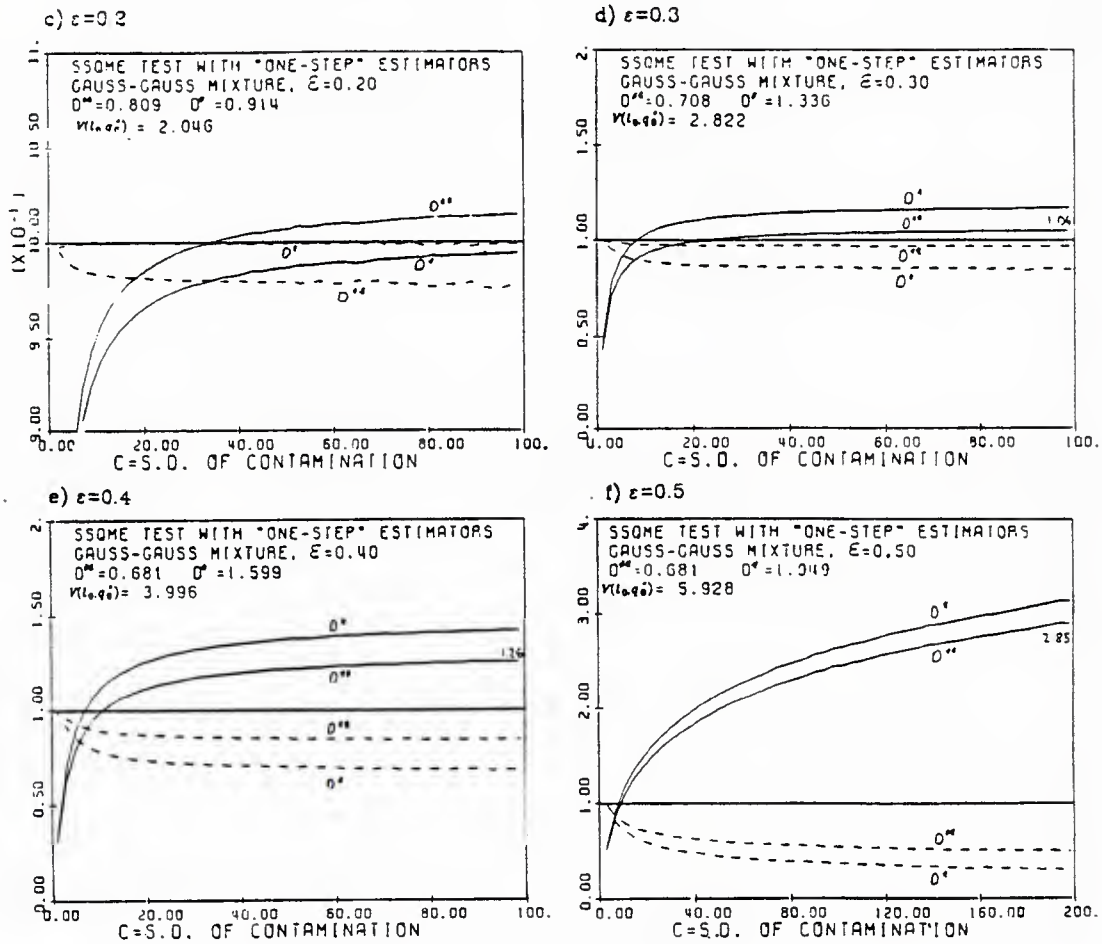


Fig. 5.3 Normalized variance $V'(F,T^{(1)})/V^{**}$ (dashed lines) and effective SNR loss $L(F,D)$ of SSQME test with a preliminary scale estimate. Gauss- Gauss mixture, c^2=contamination variance. D^* - scale normalization of the known-σ least favorable density, D^{**} - optimal scale normalization. a) ϵ=0.01 .

Fig. 5.3 (continued)



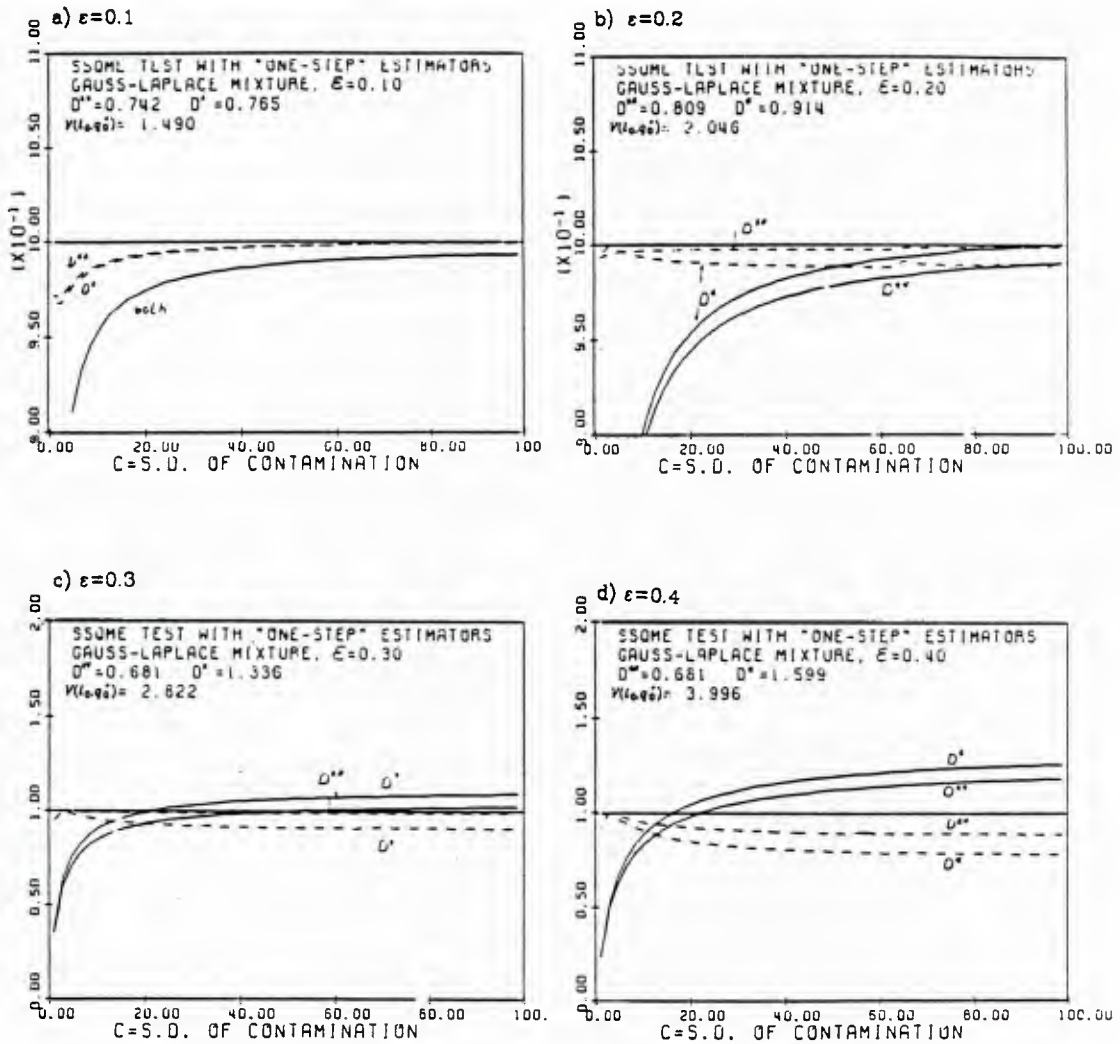


Fig. 5.4 Normalized variance $V'(F, T^{(1)})/V^{**}$ (dashed lines) and effective SNR loss $L(F, D)$ of SSQME test with a preliminary scale estimate. Gauss- Laplace mixture, c^2=contamination variance. D^* - scale normalization of the known-σ least favorable density, D^{**} - optimal scale normalization.

Finally, we note that the undesired result where the false alarm probability still depends on $f \in \mathbf{P}_0$ (unlike the SW-SSQME version) is a consequence of estimating the scale of the input samples and utilizing it for the threshold, instead of direct estimation of the variance of the location estimator. (Recall that in the classical normal case the variance of the sample mean estimator differs only by a factor of n from the variance of the input samples). This might be resolved with additional implementation complexity, by taking the adaptive threshold to be proportional to $\hat{V}^2(\mathbf{x})+\hat{V}^2(\mathbf{y})$, where

$$\hat{V}^2(\mathbf{x}) \;=\; \frac{1}{n}\left[\frac{1}{n}\sum_{i=1}^{n} IC^2\left(\frac{x_i - T_n^{(1)}(\mathbf{x})}{S_n^{(0)}(\mathbf{x})}\;,\; F_n(\mathbf{x}),\; T_n^{(1)}\right)\right] \tag{5.27}$$

and the IC is given by Eq. (5.13). Under the previously stated regularity and symmetry conditions this estimate of $V(F,T^{(1)})$ is consistent. Thus it is reasonable to assume that $V'(F,T^{(1)})$ will cancel out from the error probabilities. However, the error probabilities with this estimate are analytically intractable even when $n \rightarrow \infty$ (see also [3], Section 6.8).

We feel that this additional complication is probably unnecessary, as is evident from Figs. 5.3-5.4. Moreover, Monte-Carlo finite sample simulation results indicate that when the test of Eq.(5.16) is implemented with the breakpoints corresponding to the density q_0^*, both error probabilities did not deviate in any noticeable way from the known σ case (i.e., when (5.12) is computed with $S(F)=1$), when $\epsilon \leq 0.2$ and $n \geq 48$.

5.3 Unknown Signal Frequency and Implementation Complexity

In radar-sonar detection, the frequency of the signal is often unknown, since it is proportional to the target radial velocity by the Doppler shift effect. In the Gaussian noise case, the problem is commonly solved by employing a bank of M contiguous coherent envelope tests , each of the form of Eq.(2.4), where the i^{th} is matched (by baseband conversion) to a Doppler shift of if_r/M ($f_r = 1/T$ =pulse repetition frequency). If the Doppler shift is spread uniformly over $[0,f_r]$, almost optimal performance over this range is obtained by implementing the contiguous bank with a weighted FFT processor of length n, as the envelope of the FFT filters outputs is almost flat.[1] This will serve in the following as a basis for comparison of implementation complexity of the various robust tests.

In a similar manner, robust tests for unknown frequency can be constructed as a bank of M contiguous SSQME tests. Specifically, in the k^{th} channel, k=0,1,...M–1, the input observables of Eqs.(2.5-2.6) are transformed according to

$$I_i(k)=Re\,[(x_i+jy_i)\exp(-j\,2\pi ik/M)]\quad,\quad Q_i(k)=Im\,[(x_i+jy_i)\exp(-j\,2\pi ik/M)]$$

and then M parallel SSQME tests are performed on {I(k),Q(k)}, see Fig. A. In this way, for the M discrete frequencies kf_r/M in the uncertainty range $[0,f_r]$, the performance will be optimal as before, with some loss for straddling frequencies. The question now arises if this loss is comparable to that of the FFT processor when $M=n$. At first glance we might suspect that it would be larger due to the nonlinear processing, but it turns to be identical at least for large sample sizes.

Proposition 5.3 Let the frequency shift of the signal be $f_d=\Delta f_r/n$ (Δ measures the frequency deviation from the nearest filter, normalized by the filter bandwidth) and let $n\rightarrow\infty$. Then, the detection probability of the various SSQME tests is given by the same expressions as for $\Delta = 0$, where the effective SNR is attenuated by $G(\Delta)$ which is

[1]Flatness of about 1 dB is achieved with Hamming weighting.

identical to that of the FFT processor:

$$G(\Delta) \quad = \quad \left[\frac{\sin(\pi\Delta)}{\pi\Delta}\right]^2 \tag{5.28}$$

The proposition is proved in appendix D; the key steps are showing asymptotic normality for M-estimators of a sequence of r.v.'s with unequal means, and that the I and Q estimators are asymptotically jointly Gaussian and independent, even though I_i is in general not independent of Q_i. The implications should be clear: not only that it is not necessary to construct a bank of more than n contiguous SSQME tests, but it is also possible to apply the same weighting techniques that are employed with the FFT processor to get a flatter performance over the frequency range (at a price of reduced resolution and reduced SNR at $f_d = 0$). (See appendix D for the details.)

It was not possible to prove a similar result for the test implemented with the α–trimmed means. One of the difficulties is that, apparently, there is not any general result in the statistical literature on the distribution of L-statistics when the samples are not identically distributed. However, the simulation results of Chap 5.4 indicate that the above proposition is essentially valid. To support it, Fig. B depicts the frequency response $G_\alpha(\Delta)$, when no noise is present; this can be viewed as an approximation for $SNR \to \infty$. $G_\alpha(\Delta)$ was numerically computed as the average (with respect to the uniform phase) of the test statistic

$$G_\alpha(\Delta) = E_\phi\{\overline{I_\alpha}^2(\mathbf{x}, \mathbf{y}; \phi, \Delta) + \overline{Q_\alpha}^2(\mathbf{x}, \mathbf{y}; \phi, \Delta)\}\Big|_{\text{noise} = 0}$$

with $n = 16$ (the dependency on n was negligible above it). It is seen that the SNR attenuation decreases with more trimming, and is always less than that of the linear quadrature detector. From the simulations results, the presence of noise seems to compensate for this superior behavior.

Implementation complexity

The major increase in complexity stems from the nonlinear processing of the

SSQME test. Thus, the number of operations required for a single frequency must be multiplied by n resulting in $O(n^2)$ for the full range, compared to $n \log n$ for the linear FFT processor (in the following, $\log n = \log_2 n$). It does not seem possible to perform some of the nonlinear operations *before* the n frequency conversions.

For purposes of simplicity, we assume that multiplications, additions and comparisons are equally costly. Hence, the FFT processor requires $O(5n \log n + 4n)$ real operations. The SSQME test for the nominally Gaussian noise is based on α–trimmed means. If ordering the data is done by the QUICKSORT algorithm [39], the expected number of operations for an i.i.d. sequence is $O(n \log n)$ [2]. Hence α– trimming requires $O(n \log n + c(\alpha), 1 \leq c(\alpha) \leq n$. With it available, the MAD scale estimator requires $O(n \log n + 2n)$ operations in addition, and the 1-step estimator $T^{(1)}$ for the non-Gaussian nominal density needs $O(11n)$ more, assuming that the evaluation of a nonlinear function is done by interpolation between two stored values. With these, the order of complexity for the various tests of the previous section is easily obtained by counting. For the SW version we do not count the operations in the adjacent M reference cells, as the SW is performed by sequentially repeating the algorithm on adjacent spatial cells and it is only required to keep in memory the outcomes of operations that have already been performed. Table 5.4 summarizes the order of the number of operations. The last entry belongs to the narrow-band Wilcoxon nonparametric test [40] which is a candidate for comparison with the robust tests. The test statistic for a single frequency is

$$R_{NBW}(\mathbf{x},\mathbf{y}) = [\sum_{i=1}^{n} R^{+}(x_i) U(x_i)]^2 + [\sum_{i=1}^{n} R^{+}(y_i) U(y_i)]^2 \qquad (5.29)$$

where $R^{+}(x_i)$ is the rank of $|x_i|$ in $|x_1|, |x_2|, \ldots |x_n|$, and $U(x_i)$ is the unity step function. The test has the same order of complexity as the SSQME test, which substantially outperforms it as will be shown in the next section.

[2] QUICKSORT is a random algorithm, it can require up to $O(n^2)$ operations.

Type of Test	Order of Complexity
FFT processor	$5n\log n + 4n$
SW-SSQME for nominal Gaussian noise	$2n^2\log n + 2n\,(c\,(\alpha)+1)$
SSQME with preliminary scale estimate for nominal Gaussian noise	$4n^2(\log n + 1) + 2n\,(c\,(\alpha)+2)$
SW-SSQME for nominal non-Gaussian noise	$4n^2(\log n + 12) + 8n$
SSQME with preliminary scale estimate for nominal non-Gaussian noise	$4n^2(\log n + 12) + 11n$
Wilcoxon NB	$2n^2(\log n + 3) + 4n$

Table 5.4*a* Order of complexity for various tests of narrow-band input of length n.

As an example for the order of the actual required processing rates, consider the following radar case. With a SW test and α=0.25, the number of operations per frequency channel and range bin is $O_f\,(n)=n\,(2\log n+7)+4$. The total number for the entire range domain is $O_f\,(n)T/\tau$, where τ is the width of the range bin and $T=1/f_r$ is the pulse repetition interval. The processing rate per frequency channel is the above divided by the dwell time at each spatial direction, nT. Thus, it equals $O_f\,(n)/(n\,\tau)$. Table 5.4b compares the processing rates, in million operations per second (MOPS), for $\tau=1\mu s$ (corresponding to range resolution of 150 meters), for the robust SW and FFT detectors. The robust detector requires a factor of 10 to 40 more than that required by the FFT processor, and the actual rates are higher than possible with current technology. However, they are within the projected capability of VLSI and VHSIC technologies for the near future. Moreover, the "bank of Doppler test" structure is convenient for parallel implementation, thus only the numbers at the left part of the table need be considered.

	MOPS per channel		total MOPS	
n	robust	FFT	robust	FFT
16	15.2	1.50	244	24
32	17.1	.91	548	29.1
64	19.1	.53	1220	33.9

Table 5.4b Processing rates for a radar example.

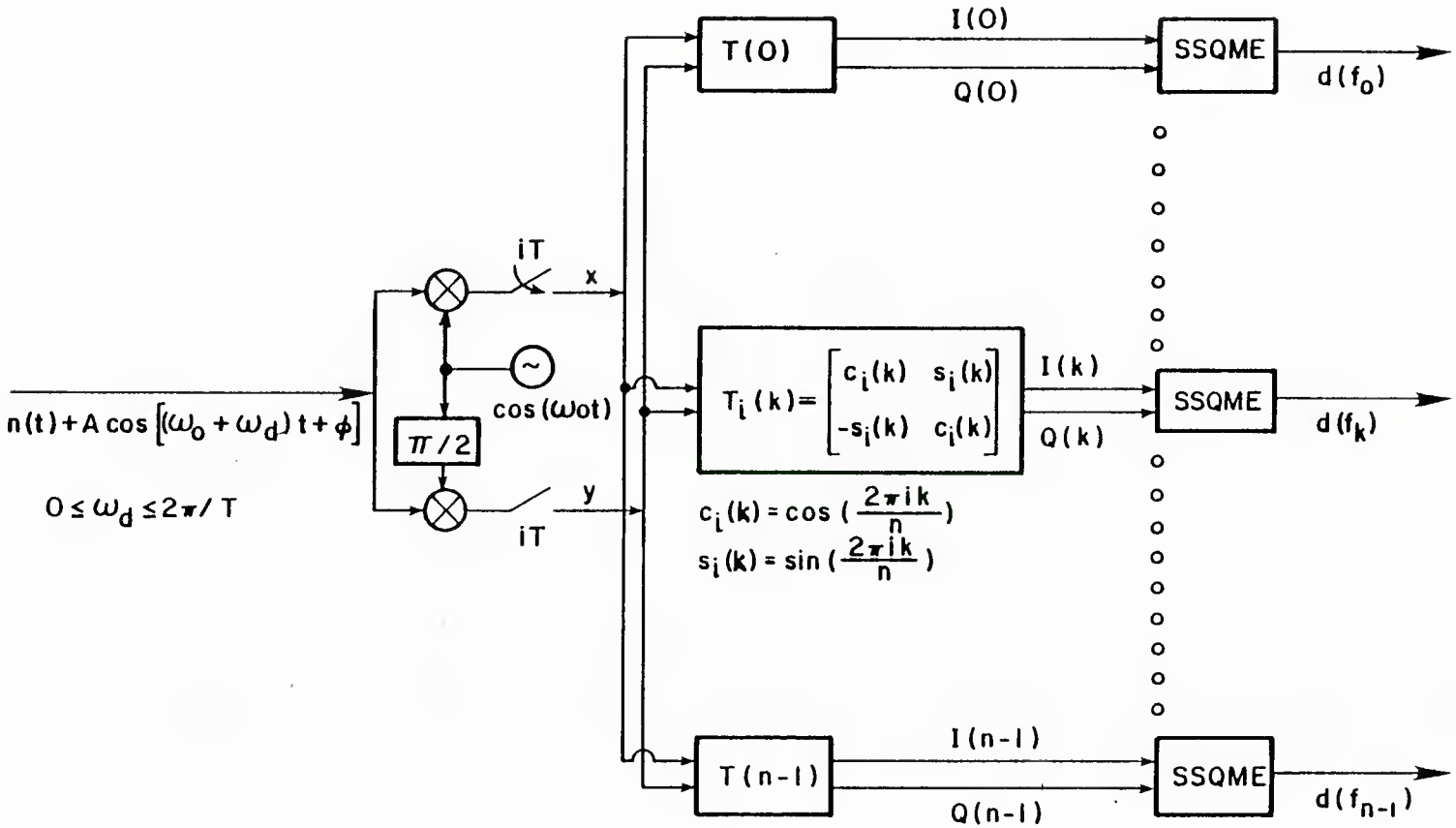
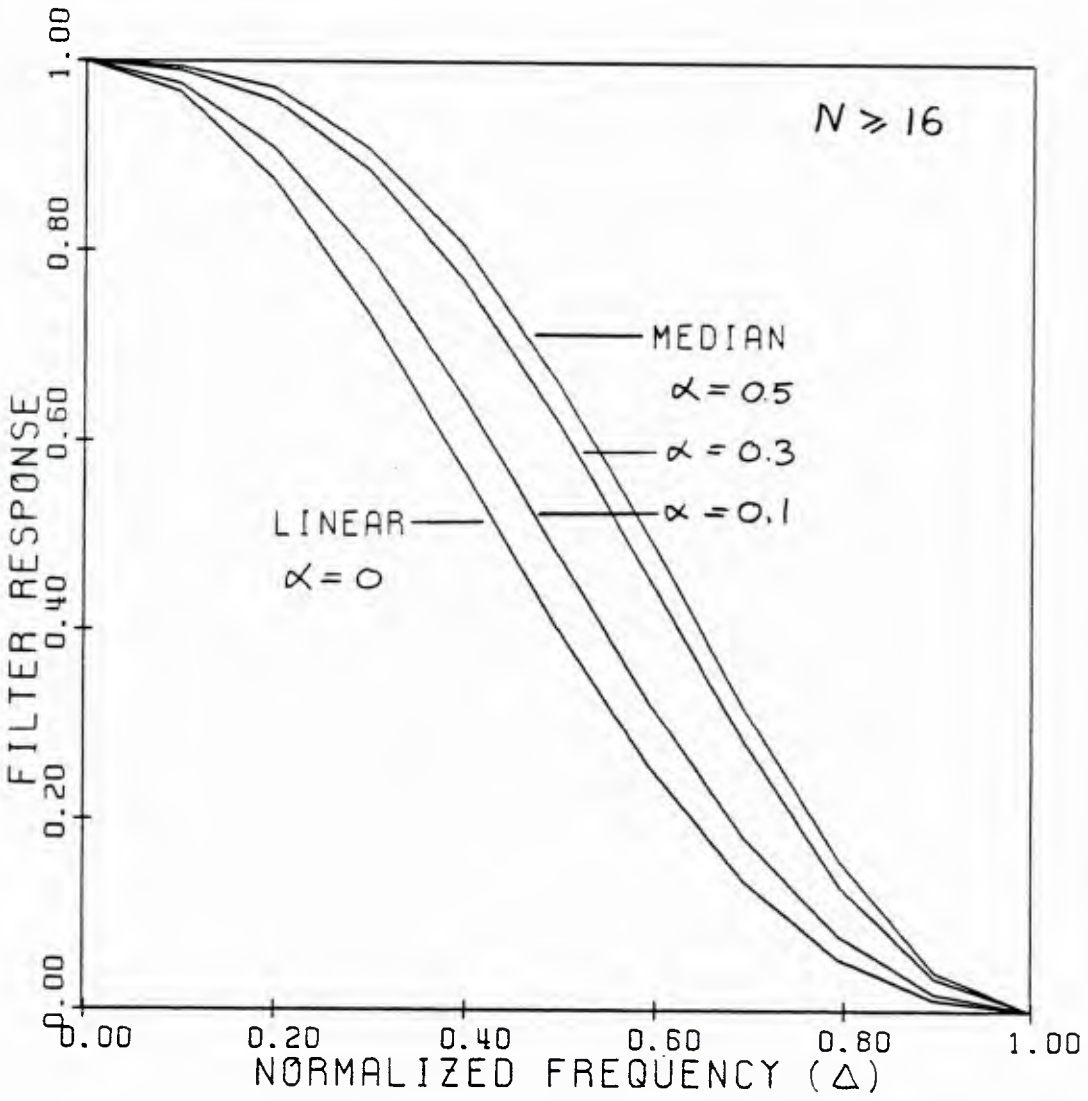


Figure *A* Bank of n SSQME tests for unknown Doppler frequency.

Figure B Frequency response of the SW test with α-trimmed estimators.



5.4 Finite Sample Simulation Results

The asymptotic approach to robust statistical procedures is an unavoidable necessity, as the $n \rightarrow \infty$ assumption allows one to invoke the central limit theorem in order to obtain *explicit* functionals which parametrize the performance of the procedure. Thus, optimization by some criterion becomes possible. Although asymptotically one could find true optimal procedures by estimating the underlying distribution, this usually requires a prohibitive number of samples [49]. Thus, the central reason for the asymptotic formulation is the hope that robust estimation and testing procedures will approach their asymptotic behavior quickly. The "Princeton Monte Carlo Study" of Andrews et al. [35] showed that this is indeed the case for the variance of M and other robust estimators of location. For detection, we need also to demonstrate the tail behavior.

In most communication engineering applications, n will be quite small. This is true particularly for radar - sonar systems, where such considerations as the desired non -ambiguous detection range, angular resolution and minimization of the required time to search the detection spatial sector, call for decreasing n. Moreover, we have seen time and again through this work that detectability of *coherent* signals is governed (at least asymptotically) by the integrated SNR nA^2. In active systems, this product is proportional to the *average* transmitter power, which usually is the fundamental constraint rather than the peak power A^2. Therefore, at a given time period the same detectability can be maintained by increasing A and reducing n while keeping nA^2 constant, thus achieving the other important systems objectives. As a result of that, most radar systems are designed with $n = O(10\text{–}50)$ or even smaller for long range systems. Consequently, the usefulness of procedures resulting from asymptotic analysis remains questionable without small - sample performance analysis.

Unfortunately, analytic tools for studying finite sample performance are not available. Results in the spirit of the Berry-Essen theorem [37], which bounds the deviation from Gaussianity of the distribution of the sample mean estimator are non-existent for

M-estimators. This is so because it is not possible to obtain an analytic expression for the finite sample moments. (Notice that its asymptotic variance was derived indirectly through the convergence of its distribution). We must resort to either numerical or simulation studies which do not analytically exhibit the rate of convergence.

For M-estimators with a monotone nonlinearity l, the distribution can be computed from an n-fold numerical convolution since $Prob\{\hat{A} < a\} = Prob\{\sum l(x_i - a) < 0\}$. For our SSQME test more integrations are required since $R = \hat{A}_x^2 + \hat{A}_y^2$, and A and ϕ have to be averaged out. Field and Hampell [42] recently showed how to alleviate the enormous amount of computations by their so called "small sample asymptotics" technique, which is closely related to Daniel's [43] saddlepoint approximation. Though this technique was shown to be very accurate down to sample sizes of 3, even in the extreme tails, it is still numerical in nature and involves a substantial programming effort. Therefore we have not proceeded in this direction but performed instead a Monte-Carlo study - programming is relatively straightforward and flexible, and the empirical distribution estimates are known to be unbiased and converge to the true distribution with a large number of repetitions.

Heuristically, the SSQME tests based on robust amplitude estimation should converge at a faster rate than the locally optimal robust detector (Section 4.4) based on weak signal assumptions. The latter incorporates a hard "non tracking" nonlinearity; the distribution of the test statistic has finite support (when n is finite) with a point mass at the upper end of the interval. In contrast, the M-estimator can take any value on R^1 and its distribution will be smoother and hence converge faster to Gaussian. This is most significant under H_1 when the distribution of the LORD test statistic will be mostly concentrated around the boundary of the intervals for large SNR.

When adopting the asymptotic theory testing structure for finite sample size, the amount of trimming or limiting must be increased if very small P_{fa} is desired, since we are dealing with the tail of the distribution where deviations from theory are larger. This

is explained more precisely as follows. Let A_k represent the event that *max { number of outliers in the I observations, number of outliers in the Q observations }* $= k$. Since these events are mutually exclusive and their union is the certain event we can write

$$P_{fa} = Pr\{R \geq t\} = \sum_{k=0}^{2\lfloor n\alpha \rfloor} Pr\{R \geq t \mid A_k\} P(A_k) + \sum_{k=2\lfloor n\alpha \rfloor+1}^{n} Pr\{R \geq t \mid A_k\} P(A_k) \tag{5.30}$$

Thus, a lower bound for P_{fa} is the second term. In it, the number of outliers exceeds the trimming capability of either estimator; hence, if t is fixed according to asymptotic theory $Pr\{R \geq t \mid A_k\}$ will be large and close to 1, and approximately $P_{fa} \cong \sum_{k \geq 2\lfloor n\alpha \rfloor+1} P(A_k)$. Assuming that the contaminated samples come from a switching model, the probability of having k outliers out of n in either channel is binomial: $P_k = \binom{n}{k} (1-\epsilon)^{n-k} \epsilon^k$, and due to independence of the quadrature samples

$$P(A_k) = P_k^2 + 2P_k \sum_{l=0}^{k-1} P_l \tag{5.31}$$

The following table exhibits this approximation for $n=16$ and $\epsilon=0.1$. With the asymptotically optimal $\alpha=0.164$, P_{fa} can not be smaller than roughly 10^{-3}, while $\alpha=0.27$ is sufficient for 10^{-6}; simulation results have validated this approximation.

α	.125	.1875	.25	.3125	.375
$k_m = 2\lfloor n\alpha \rfloor + 1$	5	7	9	11	13
$\sum_{k=k_m}^{n} P(A_k)$	$3.37 \cdot 10^{-2}$	$1.01 \cdot 10^{-3}$	$1.18 \cdot 10^{-5}$	$5.41 \cdot 10^{-8}$	$8.36 \cdot 10^{-11}$

Several versions of scale-invariant detectors of Section 5.2 were studied, for the nominally Gaussian case. They will be denoted D1 - D5 as follows:

(D1) SW-SSQME test of Eq.(5.5) based on α– trimmed estimators with the number of reference cells $M=4$.

(D2) like D1 but $M \rightarrow \infty$ (i.e.,the estimation of the variance of the test statistic con-

verges to the true value). For that, the reference statistics $W(\mathbf{u},\mathbf{v})$ of the D1 simulation was accumulated over all runs to produce the adaptive threshold. Thus, a huge saving was possible.

(D3) SW-SSQME test with a "one-step" location estimators of Eq.(5.15) based on the MAD scale estimate; its normalization factor D=0.632 was adjusted to get (empirically) consistency of the scale estimator $ES_{16}^{(0)}(\mathbf{x})$=1 at the Gaussian (this value is somewhat lower than predicted from the asymptotic theory). The nonlinearity was the soft limiter $l(x,-k,k)$.

(D4) 1S-SSQME test of Eqs. (5.16)-(5.17) - no reference cells - based on the same location and scale estimators of D3 .

(D5) SW version of the LORD from Section 4.4 (obtained from weak signal assumptions by replacing the M - estimators with the nonlinearity-integrator of Eq. (4.29)), $M \rightarrow \infty$.

For D1-D4, both α and k were adjusted to obtain better false-alarm control as discussed before; for D5 we chose k as the optimal for the given ϵ from the asymptotic theory.

For each test, two simulation programs were built. The first is for H_0 and its output are graphs of the false alarm probability vs. the normalized threshold, as well as the first two moments of the location and scale estimators. The second program is for H_1. It simulates the detection probability as a function of the SNR for several levels of the false alarm probability and for a Swerling I target model, i.e., coherent narrowband signal with Rayleigh distributed amplitude and uniform phase (see Section 4.3, case a). In each of the simulations, the test was subjected to four different noise p.d.f.'s from an ϵ–mixture family where the nominal is normal Gaussian:

(f_1) $\epsilon = 0$

(f_2) $\epsilon = 0.1$ and two point masses at ± 15

(f_3) $\epsilon = 0.1$ and a Gaussian contamination with variance c^2=100

(f_4) ϵ=0.01 and a Gaussian contamination with c^2=900

In the sequel, we shall use these numbers to identify the situation. The input variances are thus 1, 23.4, 10.9 and 9.99 respectively - all $\epsilon \neq 0$ cases represent large, heavy tailed contaminations and would have resulted in a substantial degradation of performance if the unrobustified envelope detector was employed (see Section 2.2). The ϵ– mixture r.v.'s were generated from a switching device: at each step, a normal r.v y_i was generated together with a binomial r.v s_i that is equal to 1 with probability ϵ; $x_i = y_i$ if $s_i \neq 1$ and $x_i = c \cdot y_i$ (or $x_i = \pm 15$ with probability 0.5 for f_2) otherwise. Almost all simulations were for n =16.

Table 5.5 displays the Monte- Carlo variances of the α– trimmed estimator for various α's. For comparison, from the asymptotic theory at the least-favorable p.d.f. $V^*(\epsilon=0.1)$ =1.49, and for α=0.225 we computed from Eq.(5.3) $V(1)$=1.17, $V(2)$=1.51, $V(3)$=1.46, $V(4)$=1.2. This good correspondence between asymptotic theory and finite sample simulation (for n=20) was observed before by Andrews et. al. [35]. The asymptotically optimal trimming ratio is $\alpha(\epsilon=0.1)$= 0.164; as was discussed before it is necessary to apply more trimming with small sample size when very small false alarm probabilities are desired. The simulations indicated that α=0.225 to 0.3 is sufficient; the table shows that except for f_1 for this α range the variances are roughly as those for α = 0.164 (or even smaller). The results suggest that the detectability loss compared to theory is small. This has been verified from the H_1 simulation results, which will be presented below. Similar conclusions can be drawn from Table 5.6, which is for the "one-step" estimator; here the asymptotically optimal limiting is $k(\epsilon=0.1)$ = 1.14, and very small losses (if any) are incurred by taking k =0.75.

α	$V(f_1)$	$V(f_2)$	$V(f_3)$	$V(f_4)$
0.164	1.09	1.43	1.60	1.13
0.225	1.15	1.34	1.53	1.18
0.25	1.17	1.35	1.51	1.20
0.3	1.21	1.41	1.55	1.25
0.4	1.36	1.56	1.69	1.39
0.5	1.47	1.68	1.81	1.50

Table 5.5 Monte-Carlo variances of the α– trimmed estimator. n=16, 20,000 runs.

k	$V(f_1)$	$V(f_2)$	$V(f_3)$	$V(f_4)$
1.14	1.08	1.29	1.53	1.12
0.9	1.13	1.31	1.51	1.16
0.75	1.16	1.34	1.50	1.24
0.5	1.29	1.46	1.62	1.32
0.3	1.41	1.60	1.74	1.42

Table 5.6 Monte-Carlo variances of the "one - step" estimator with MAD scale estimator. n=16, 20,000 runs.

Probability of false alarm curves are shown in Figures 5.5 - 5.16. In all these figures, the smooth curves marked "T" are computed from the asymptotic theory - Eqs.(5.6) and (5.23), and the numbered curves are the simulation results for the densities f_1– f_4. Fig. 5.5 corresponds to D1 and 5.6 to D2, with α=0.165. We see clearly that here α is too large, since for the more contaminated cases P_{fa} is much higher than desired. This problem is solved by taking α=0.225 as can be seen from Figs. 5.7-5.8. In the first one, the Monte-Carlo curves are all **remarkably close** to the theoretic curve, within the range where the simulation results are reliable [1], $P_{fa} \gtrsim 4/n_r = 2 \cdot 10^{-4}$. The test is clearly CFAR, for very different noise p.d.f.'s.

In Figure 5.8 P_{fa} is even somewhat smaller than the theoretic, (this means that the distribution of the test statistic R has shorter tails compared to the one sided exponential).

[1] The empirical distribution for n_r repetition of the Monte-Carlo experiment $\hat{F}_{n_r}(x)$=number$\{x_i \geq x\}/n_r$, is a binomial r.v. with mean $=1-F(x)$ and variance $=F(x)(1-F(x))/n_r$. Therefore, for $P_{fa} \ll 1$,for $\sqrt{Var(\hat{P}_{fa})}/P_{fa} \leq 0.5$ one obtains $P_{fa} \geq 4/n_r$.

Thus, if the threshold is fixed according to the asymptotic theory, P_{fa} is bounded from above as desired. The same is valid for α=0.3 as can be seen from Figs. 5.9-10. The curves for D3 are shown in Figs. 5.11-12; in the first one, the limiter breakpoint is k=1.14 according to the asymptotic theory - it is again seen that the limiting is not "hard" enough, as P_{fa} for f_3 is higher than the theoretic bound when $P_{fa}<0.01$. This problem is corrected by taking k=0.75 as can be seen from Fig. 5.12.

Results for D4 are displayed in Figs. 5.13-5.14; notice the two sets of curves in Fig. 5.13: the first is of $\log Prob\{R(\mathbf{x},\mathbf{y}) \geq (t/n) V^* W(\mathbf{x},\mathbf{y})\}$ where V^*=1.49, which should approach a straight line when $n \rightarrow \infty$ according to proposition 5.2. In the second case the scale is contracted by taking V^*=3. Here asymptotic theory is no longer valid: the curves are not linear and for achieving small P_{fa}'s higher threshold settings are required. These higher settings, in turn, would cause higher detectability losses. This deviation from asymptotic theory is not surprising, and is equivalent to the behavior of the "Quadrature t-test," which is the maximum likelihood test for a coherent narrowband signal in Gaussian noise of unknown variance

$$(\frac{1}{n}\sum x_i)^2 + (\frac{1}{n}\sum y_i)^2 \underset{H_0}{\overset{H_1}{\gtrless}} \frac{t}{n}[\sum(x_i - \overline{x})^2 + \sum(y_i - \overline{y})^2] \qquad (5.32)$$

For small n, the right hand side cannot be approximated by a Gaussian distribution (it is chi-squared *for all* n), and also the dependency between the right and left sides of Eq. (5.29) cannot be neglected as we did in proving Proposition 5.2. In Fig. 5.14 n was increased to 48, in order to check convergence. It is evident that the curves almost converge to asymptotic theory, although some curvature is still noticed.

Finally, Fig. 5.15 is for the LORD test D5. Here, as in Fig. 5.8, a tendency for shorter tails is observed. This could be heuristically described by the hard limiting that restricts the distribution of the test statistic to have a finite support.

Dramatic improvement in false alarms control over the unrobustified SW detector Eq. (2.8) is evident when comparing the previous figures with Fig. 5.16. When it operates

in a Gauss-Gauss (ϵ=0.1, c =10) noise, P_{fa} is intolerably increased from the design 10^{-4} to about 0.1 when M=4, and from 10^{-6} to $2\cdot10^{-4}$ when $M\rightarrow\infty$. Compare also Fig. 2.5 for ϵ=0.01. Thus we can confidently conclude that the various SW-SSQME tests essentially maintain CFAR according to the asymptotic theory of Section 5.2 even for n as small as 16 with P_{fa} down to (at least) 10^{-4}. For the test without reference samples D4, n is not much larger than 48. Based on these curves, the detection thresholds for the H_1 simulation were simply taken from the smooth theoretical curves, and extrapolated from them for P_{fa}=10^{-6}. Future work could utilize reduced variance sampling techniques such as "importance sampling "[44] to validate this extrapolation.

Probability of detection curves for D1, D2 and D5 are shown in Figs. 5.17-27. Here 5000 repetitions are sufficient to get $Var^{1/2}(\hat{P}_d)/P_d \leq 0.1$ for $.02 \leq P_d \leq 0.98$. The abscissa is the effective integrated $SNR = nE(A^2)/\sigma^2$ where σ^2 is the variance of the *nominal* Gaussian p.d.f. The known signal frequency case is depicted in Figs. 5.17-20. The Monte-Carlo curves are clearly bounded from the left by the computed P_d for the unrobustified SW detector when ϵ=0, and from the right by the maximin lower bound computed from asymptotic theory, Eq. (5.6), with V^* of the least-favorable p.d.f. f^*. Here again the fit between asymptotic theory and small sample performance is striking: the curves for f_1–f_4 are ranked *exactly* according to the estimation variances of Table 5.5, and the simulated P_d values generally differ by no more than a few percent when the simulation variances are substituted in Eq. (5.6) or (5.23). Equivalently, the horizontal differences between curves are roughly the dB values of the variances from Table 5.5. We also note that the difference between α=0.225 and α=0.3 is very small, again according to the variance difference.

In contrast with the above results, the unrobustified SW detector of Eq.(2.8) was found to suffer substantial detectability losses. They were almost identical to the increased *input* variances of $f_2 - f_4$, with respect to f_1.

Some non-parametric schemes were developed for detection of narrowband signals in unknown noise; the narrowband Wilcoxon detector, Eq.(5.29), was proposed by Carlyle and its Monte-Carlo performance in purely Gaussian noise was studied by Hansen [40]. It was found that the SNR losses compared to the UMP detector for n=16 and P_{fa}=10^{-6} are 5 and 14 dB, for P_d=0.5 and 0.9, respectively (smaller losses are incurred for larger n) - compared to merely 0.6 db for D2 in Fig. 5.18 ! Furthermore, the performance in non-Gaussian noise has not been studied.

The performance of the detector without reference samples, D4 with $n = 16$, is shown in Fig. 5.20. Although convergence to the asymptotic theory is not yet reached (in accordance with the higher required threshold values - recall Fig. 5.13), the detector is clearly robust. This is evident from the small differences between the curves for $f_1 - f_4$. By comparison with Fig. 5.18, the SNR losses are roughly 1.8 and 5.8 dB for $P_{fa} = 10^{-2}$ and 10^{-6}, respectively, independent of the P_d. However, when comparing it with the SW-SSQME test D1 of only 4 reference channels, D4 outperforms it by roughly 1 and 3.4 dB, for $P_{fa} = 10^{-2}$ and 10^{-6}, respectively. By virtue of the shift invariance of the one-step location estimator, if convergence to Gaussianity is reached under H_0, the performance under H_1 will agree with the asymptotic prediction. Thus, when $n \geq 50$ D4 would hardly suffer any detectability losses compared to the UMP detector for the least favorable noise p.d.f.

In contrast to the excellent performance of the structures proposed in this work, Fig. 5.21 shows that while the LORD test D5 performs roughly the same for P_{fa}=10^{-2}, it is *inferior* for P_{fa}=10^{-6}; its SNR should be higher by 4.6 dB to achieve P_d=0.9 at f_3. This can be simply explained: when the SNR is sufficiently high for high P_d, the "non-tracking" hard limiter of Eq. (4.29) prevents the distribution of R from tilting towards higher values. Also, asymptotic theory does not properly describe its behavior at the most interesting zone $P_d \rightarrow 1$. (Recall that SNR $\rightarrow 0$ was *necessary* for the derivation of its asymptotic distribution, but not for the tests which are based on true robust

estimation of the arbitrary amplitude). Notice also that the LORD test D5 is not scale invariant, and the simulation results for it are valid only when the variance of the nominal equal unity.

The performance with unmatched signal frequency when a bank of n contiguous SW-SSQME tests is utilized for unknown frequency (Section 5.3), are depicted in Figs. 5.22-24 where the normalized deviation $\Delta f \triangleq n \Delta f_d / f_r = 0.25$, and in Figs. 5.25-27 for $\Delta f = 0.5$ (halfway between the central frequency of adjacent tests). The mismatch loss is almost indistinguishable from that of Proposition 5.3 - i.e., the same loss incurred by the FFT processor for a frequency unmatched signal in pure Gaussian noise. While this loss is reasonable for D1 and D2, 0.91 dB and 3.9 dB for $\Delta f = 0.25$ and 0.5 respectively, the weak signal locally - optimal test D5 **totally breaks down** under this situation, as demonstrated by Figs. 5.24 and 5.27. We note that this loss for D_1 and D_2 can be reduced by weighting the input samples. (See appendix D.)

Another interesting conclusion can be drawn. In many practical cases, the noise p.d.f. is variance constrained, as it is proportional to the total (nominal plus contaminating components) noise power. This is true in particular for the clutter (reverberation) environment, where the noise power is proportional to the radar (sonar) transmitted power. A question now arises as to which variance constrained noise p.d.f. constitutes the worst detection environment. Consider for example Fig. 5.18 and f_3 for which the detector D2 is essentially optimal. However, the variance of f_3 is 10.9. Consequently, in a Gaussian noise environment of equal power $\sigma^2 = 10.9$, the UMP detector would perform worse by 10.9 -1.5 = 9.4 dB than the detector D2 ! This turns to be true in general. In [45], we show that the Gaussian is the least favorable variance constrained p.d.f. for detection of lowpass as well as narrowband coherent signals with random parameters.

Fig. 5.5 False alarm probability. Test - D1 , α=.164.

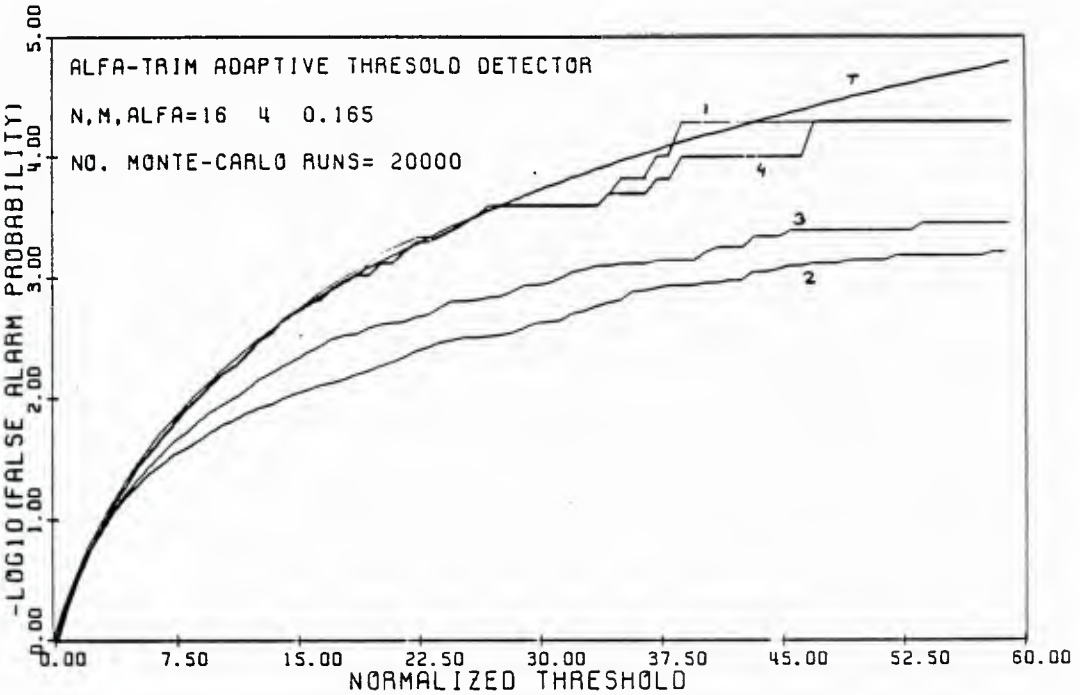


Fig. 5.6 False alarm probability. Test - D2 , α=.164.

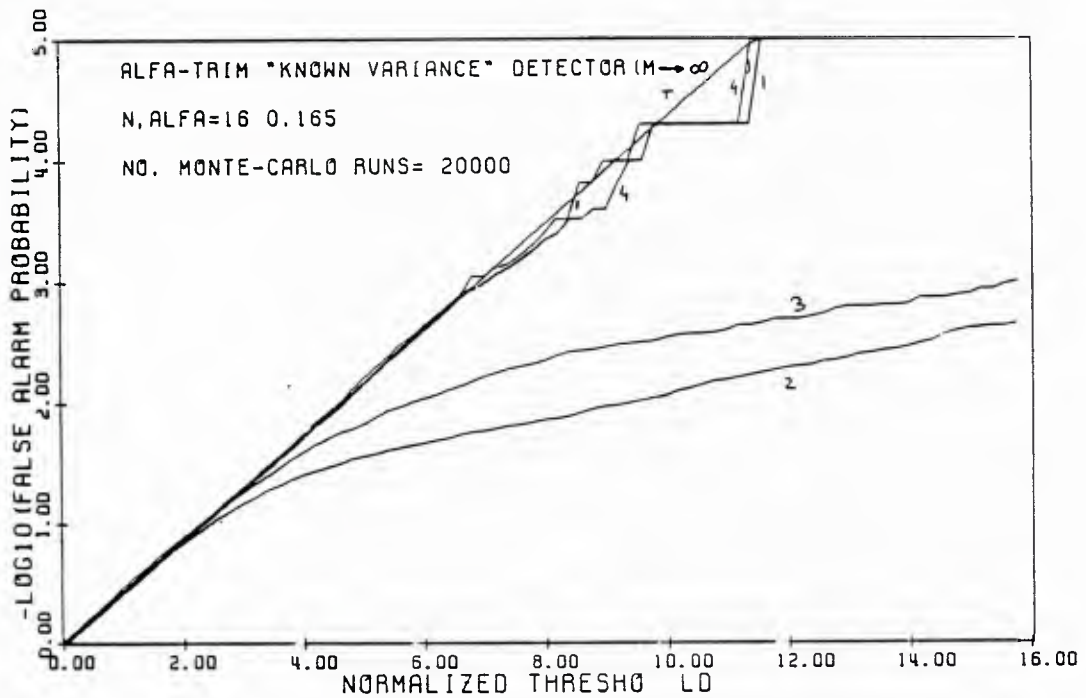


Fig. 5.7 False alarm probability. Test - D1 , α=.225.

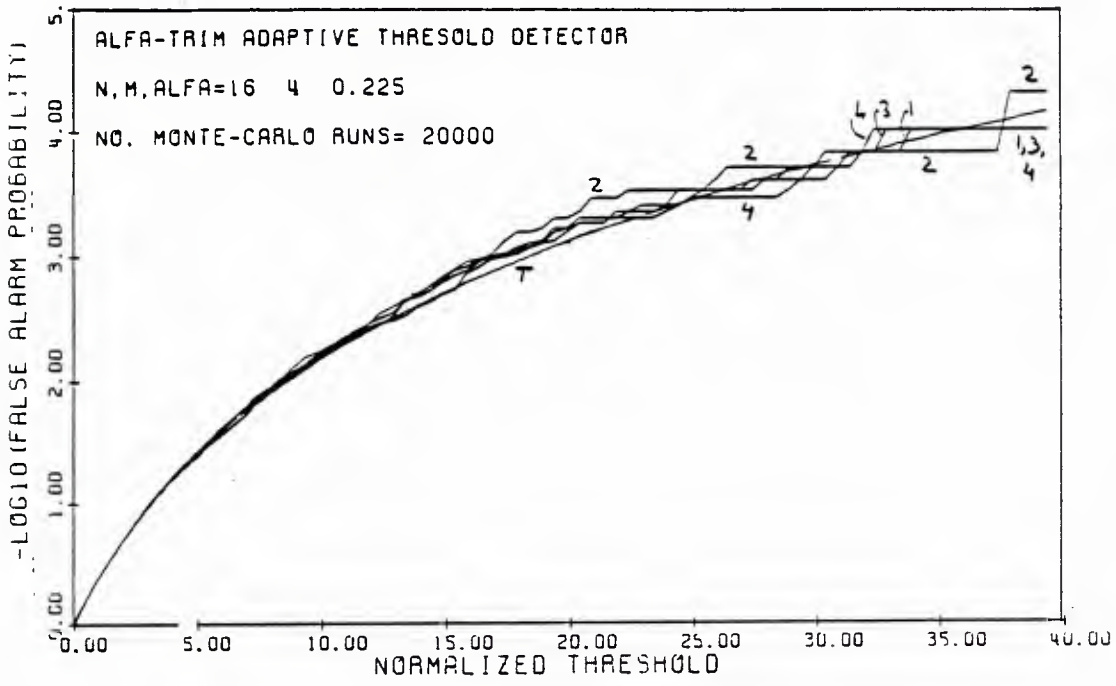


Fig. 5.8 False alarm probability. Test - D2 , α=.225.

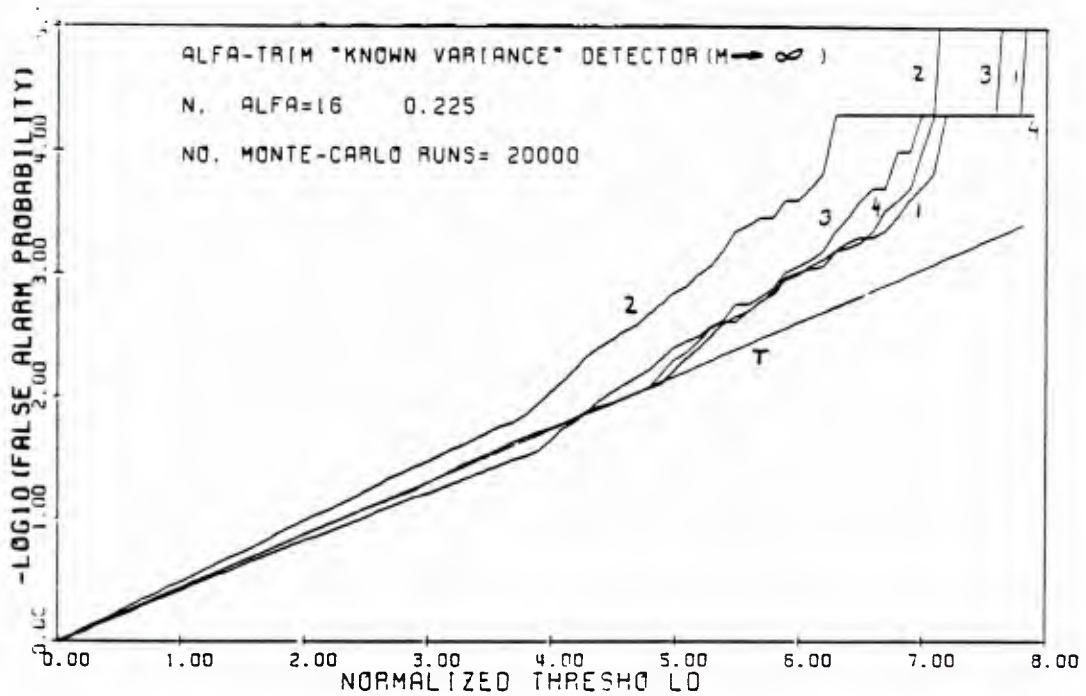


Fig. 5.9 False alarm probability. Test - D1 , α=.3.

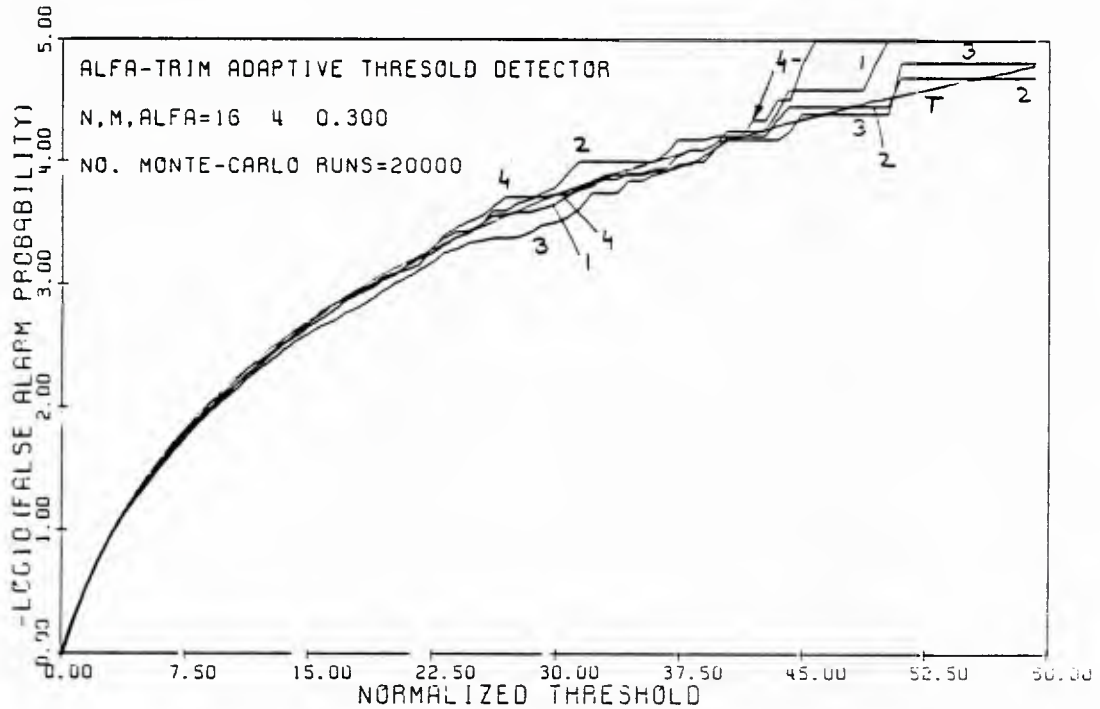


Fig. 5.10 False alarm probability. Test - D2 , α=.3.

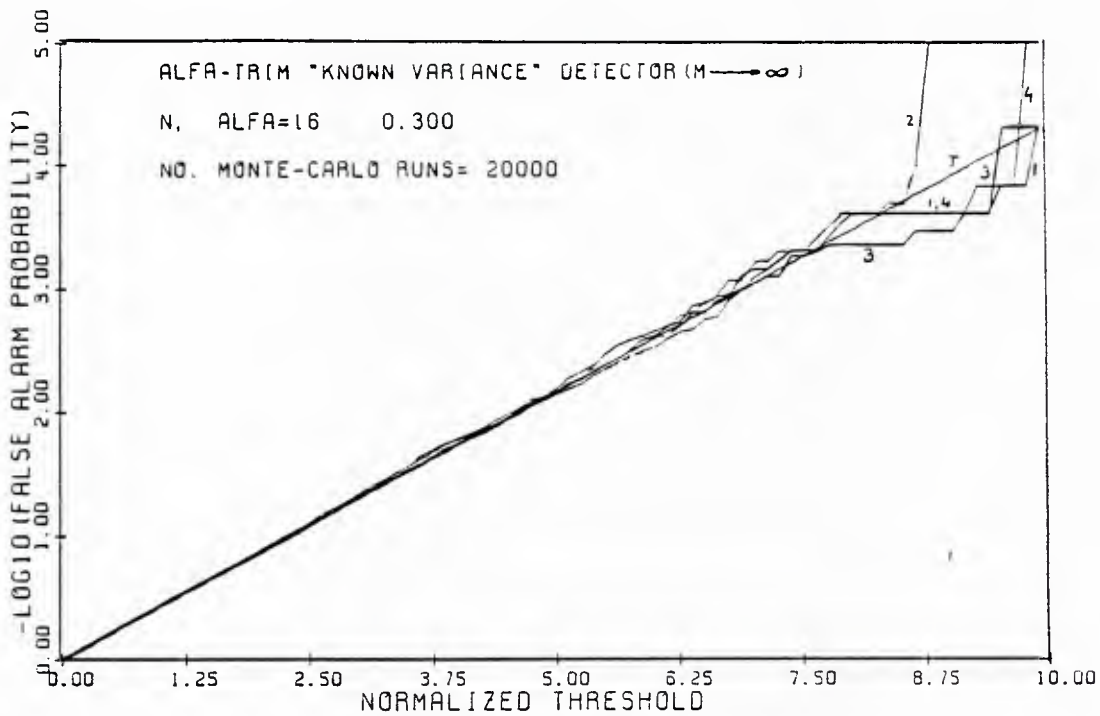


Fig. 5.11 False alarm probability. Test - D3 , K=1.14.

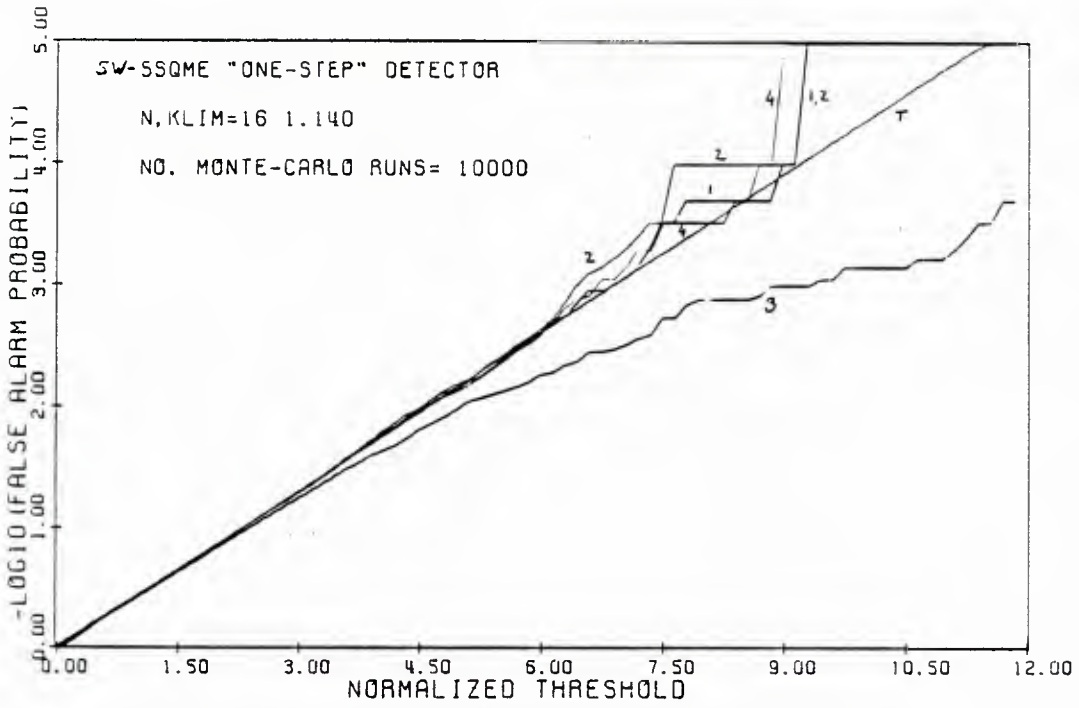


Fig. 5.12 False alarm probability. Test - D3 , K=.75.

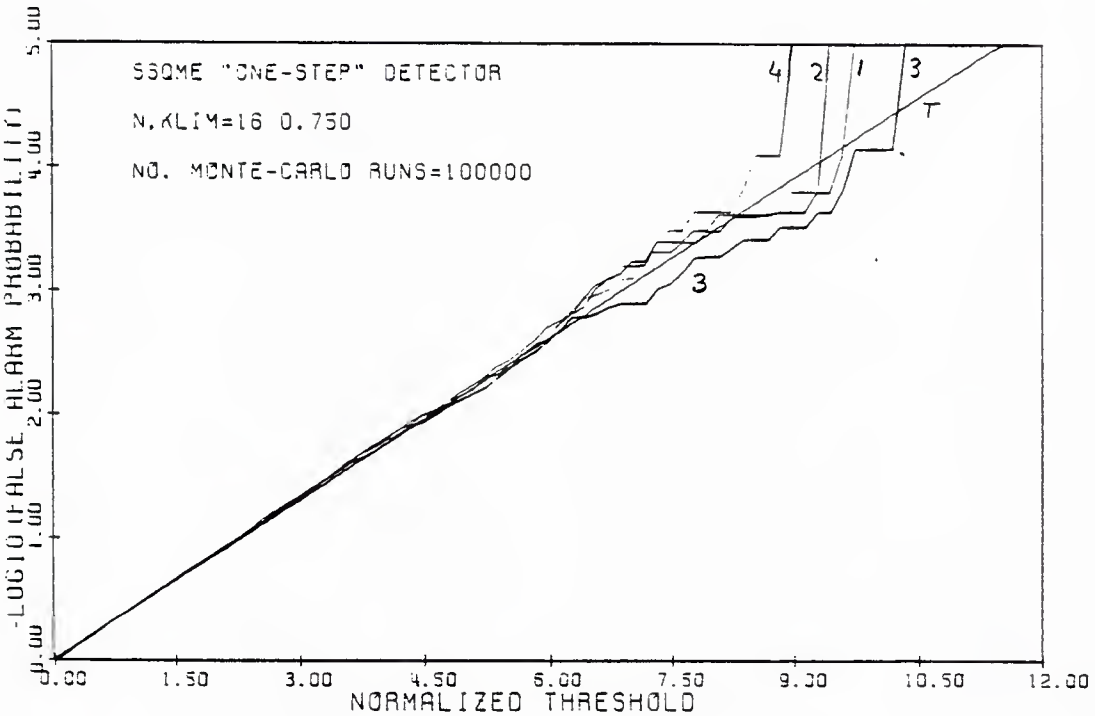


Fig. 5.13 False alarm probability. Test - D4 , K=.85, n=16.

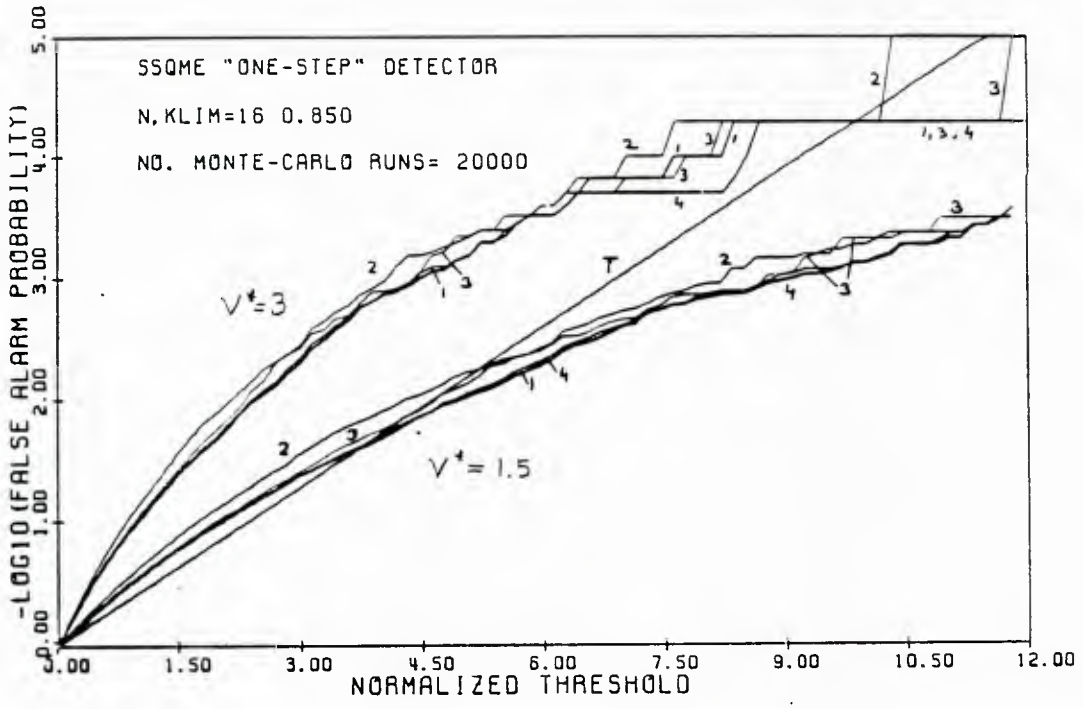


Fig. 5.14 False alarm probability. Test - D4 , K=.9, n=48.

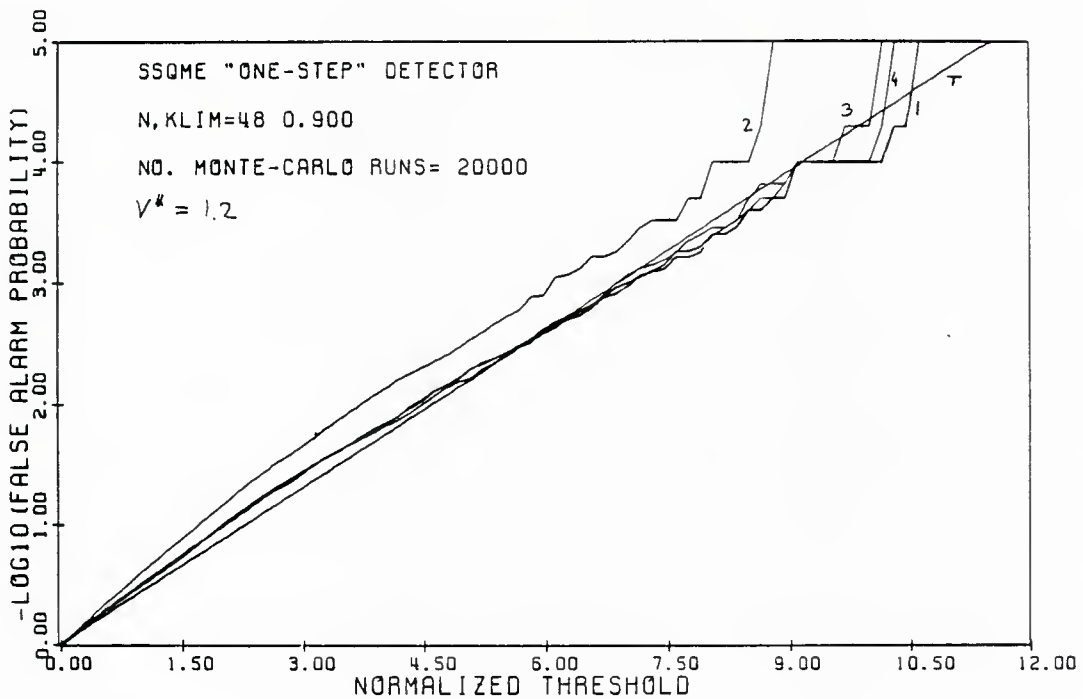


Fig. 5.15 False alarm probability. Test - D5 , K=1.14.

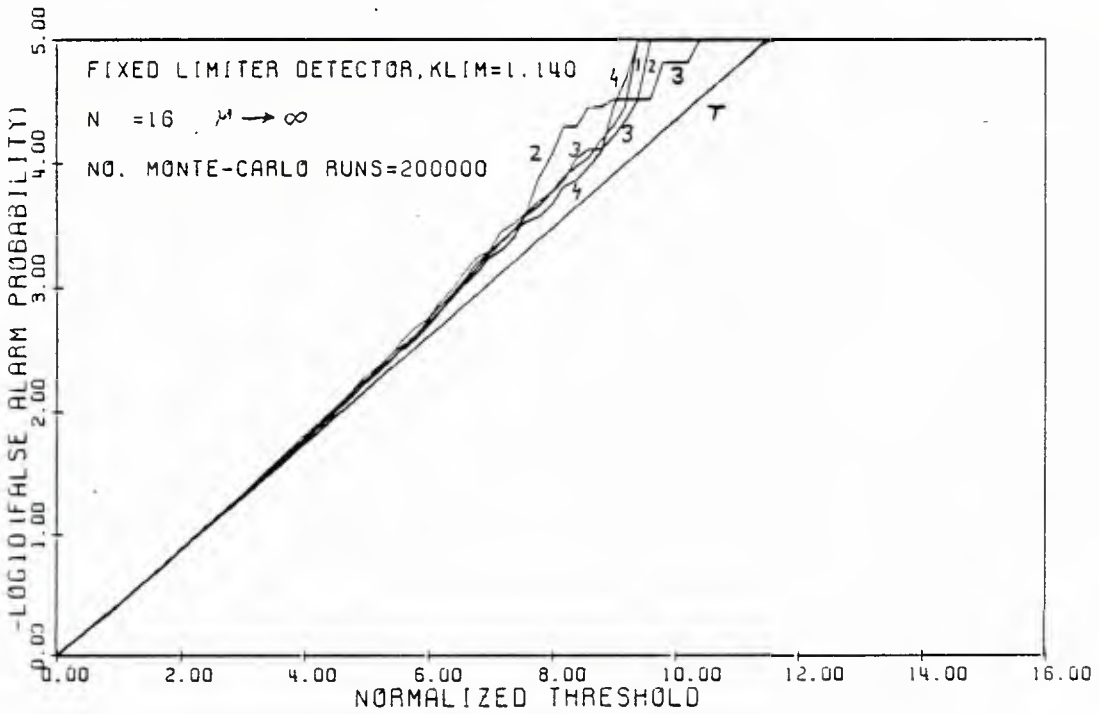


Fig. 5.16 False alarm probability. Unrobustified envelope detector, Gauss-Gauss mixture (ϵ=0.1,c =10).

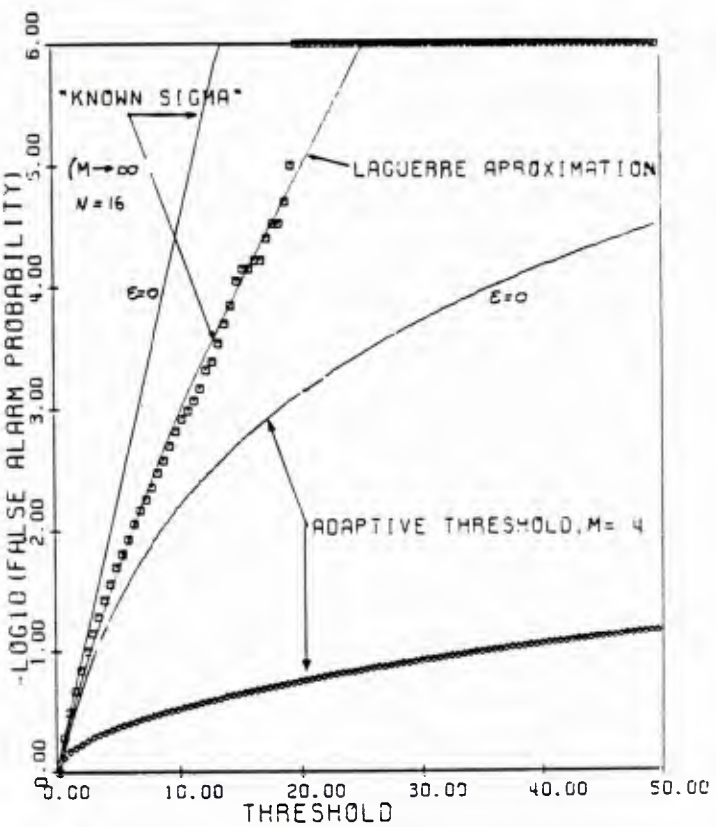


Fig. 5.17 Detection probability. Test - D1, α=.225,Δf =0

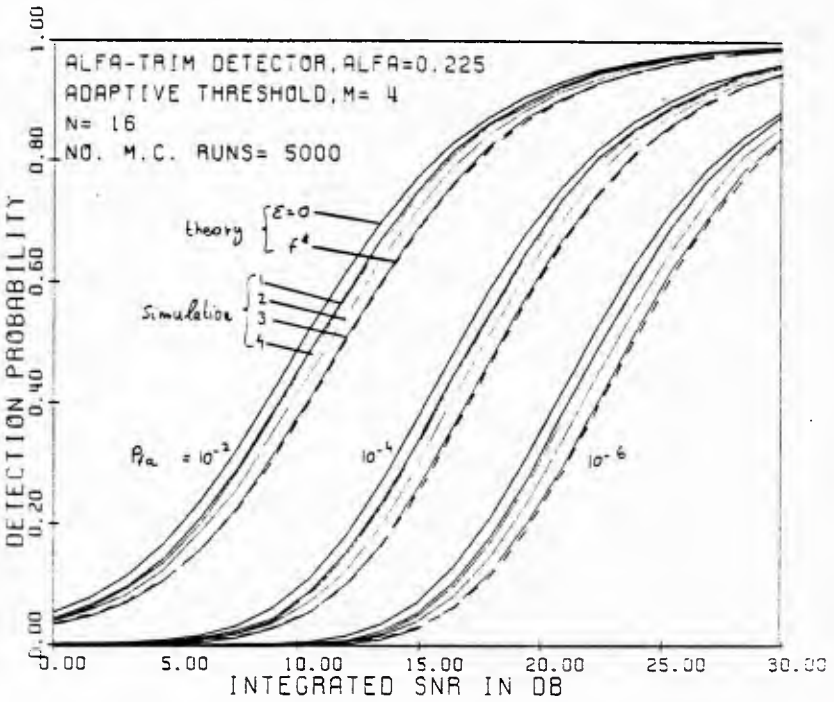


Fig. 5.18 Detection probability. Test - D2, α=.225,Δf =0

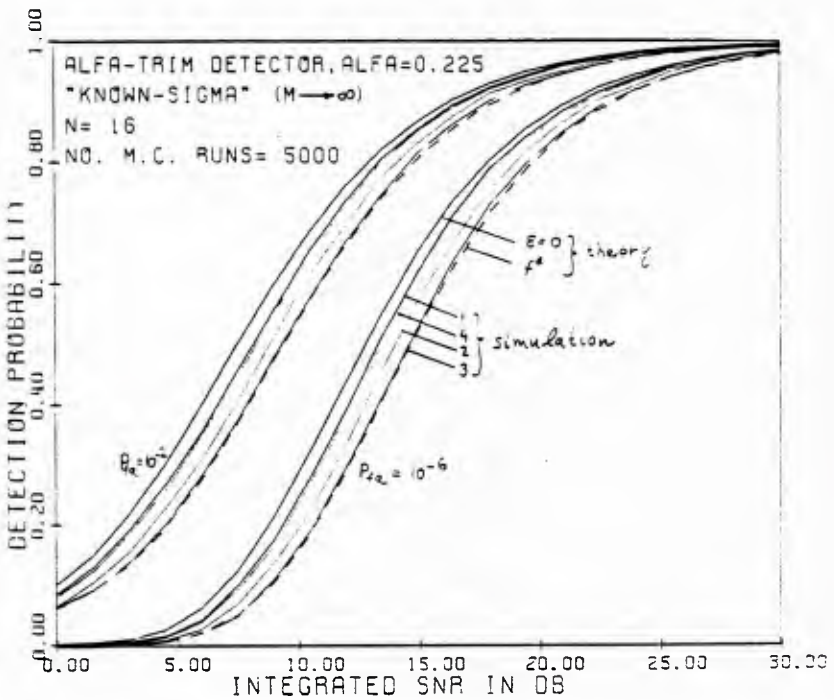


Fig. 5.19 Detection probability. Test - D1, α=.3,Δf =0

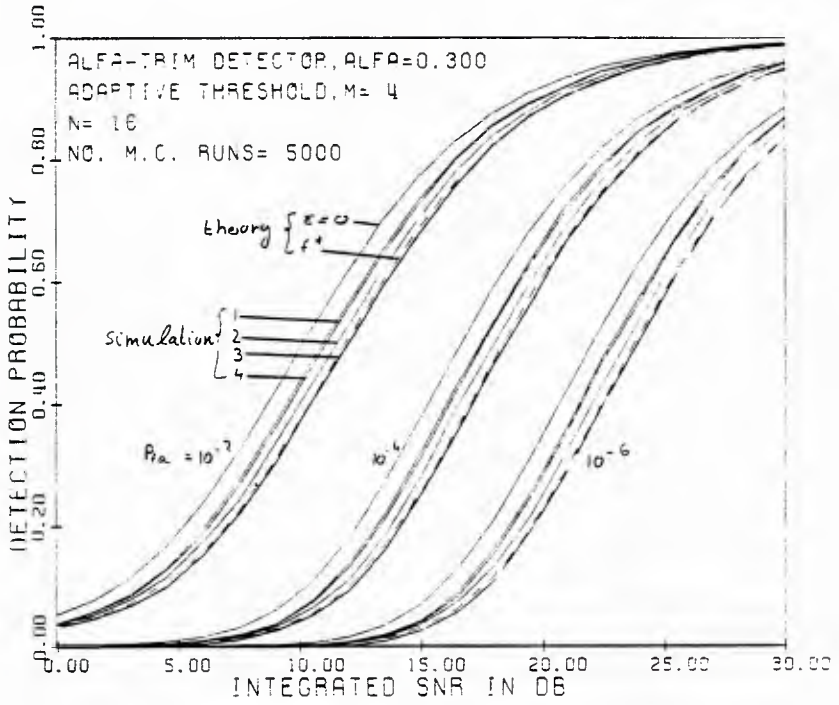


Fig. 5.20 Detection probability. Test - D4, n = 16, Δf =0

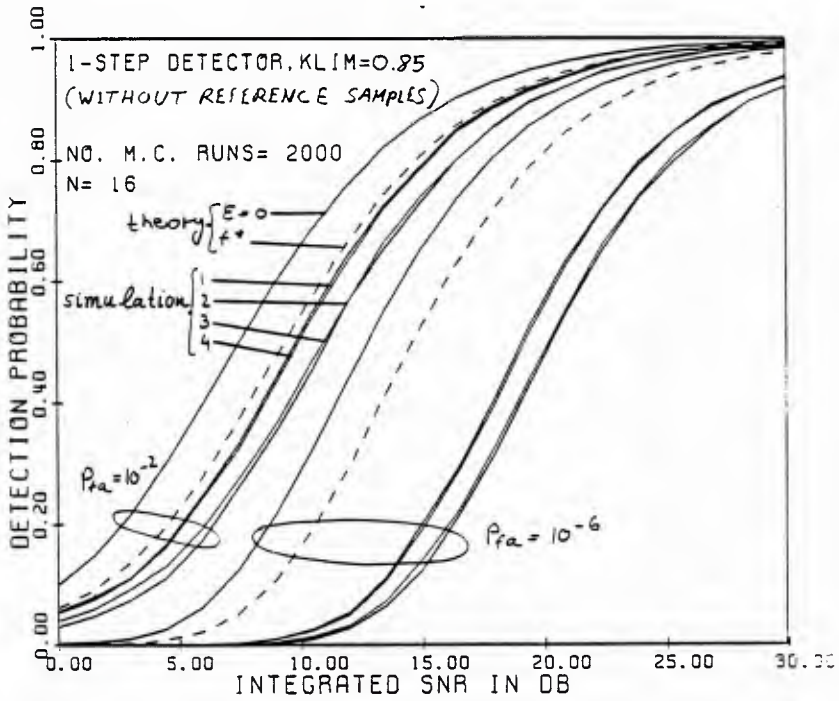


Fig. 5.21 Detection probability. Test - D5, K = 1.14, Δf =0

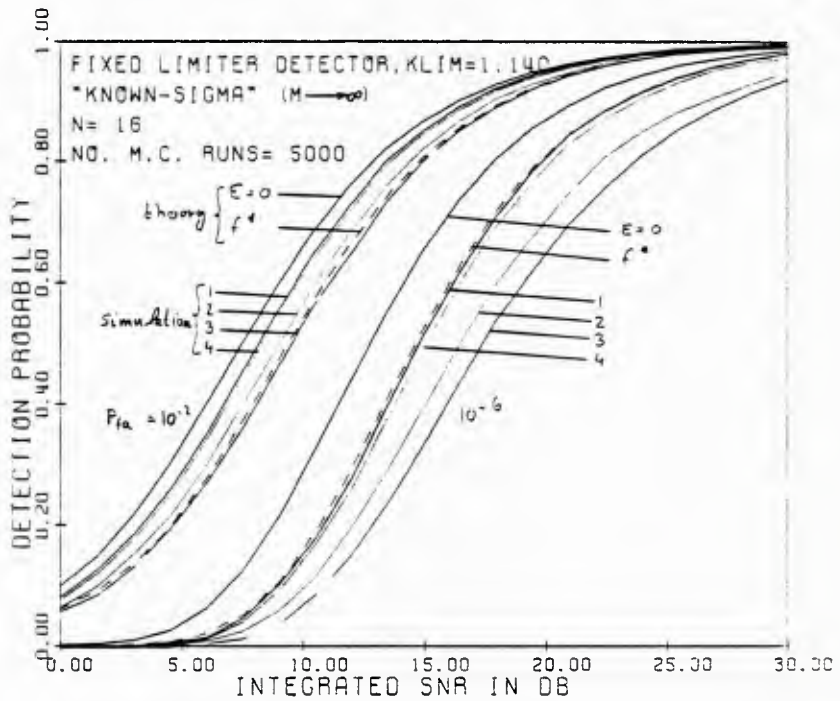


Fig. 5.22 Detection probability. Test - D1, α=.225,Δf =.25

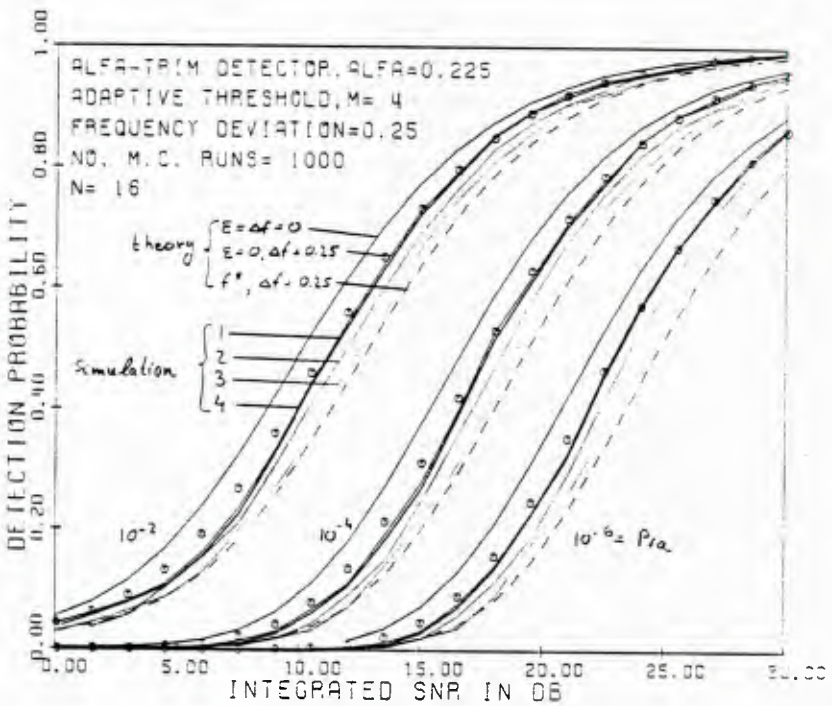


Fig. 5.23 Detection probability. Test - D2, α=.225,Δf =.25

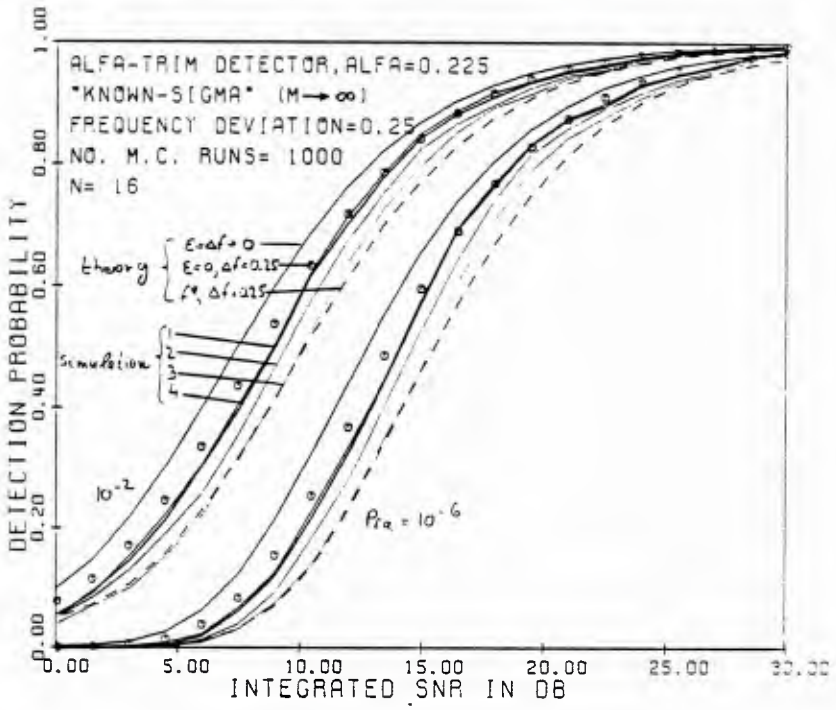


Fig. 5.24 Detection probability. Test - D5, K =1.14,Δf =.25

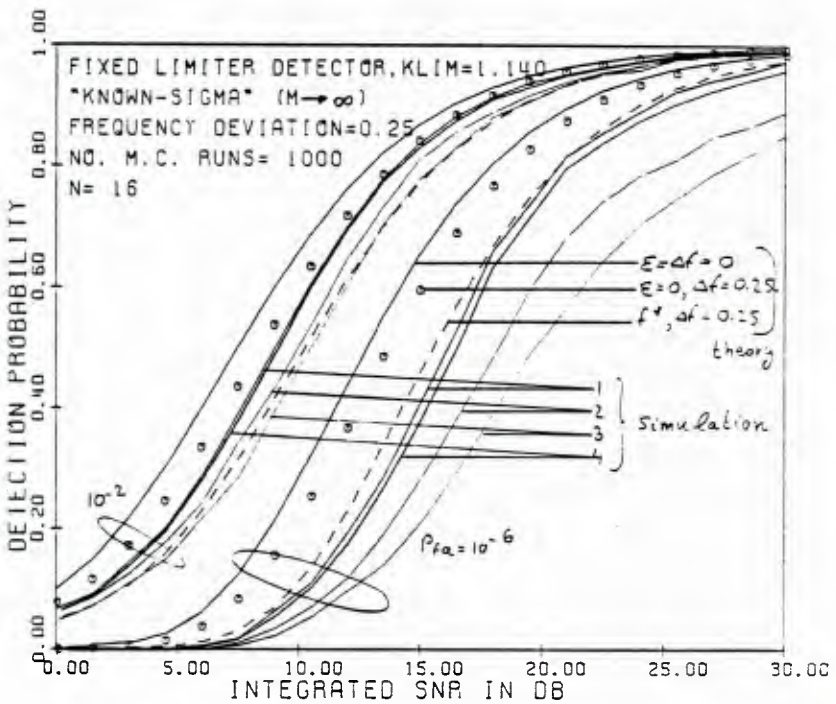


Fig. 5.25 Detection probability. Test - D1, α=.225,Δf =.5

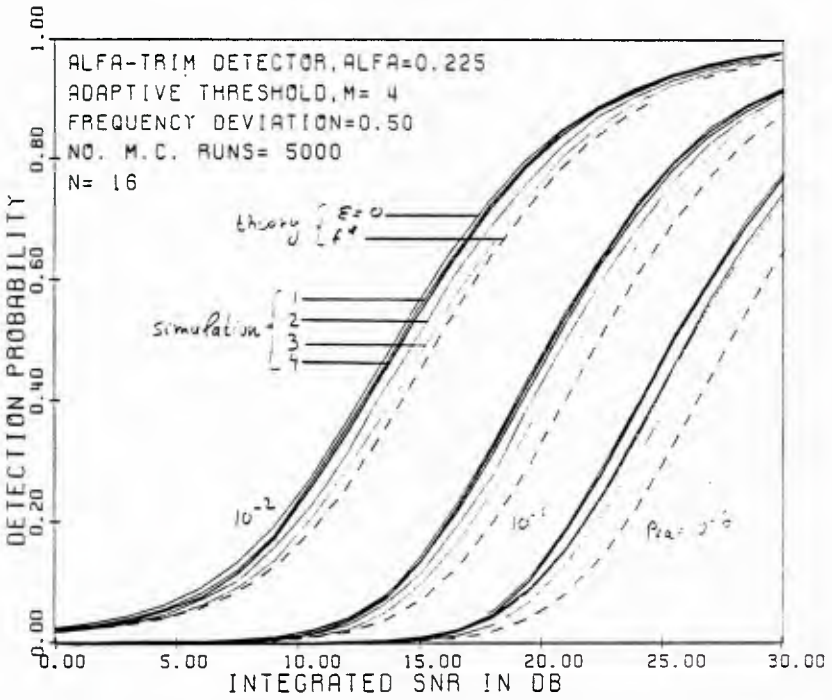


Fig. 5.26 Detection probability. Test - D2, α=.225,Δf =.5

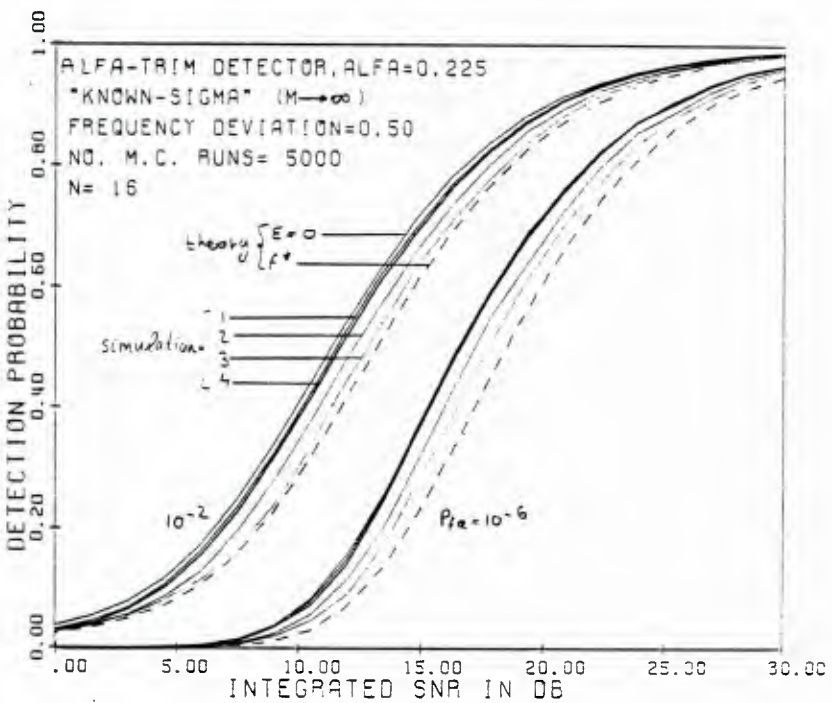
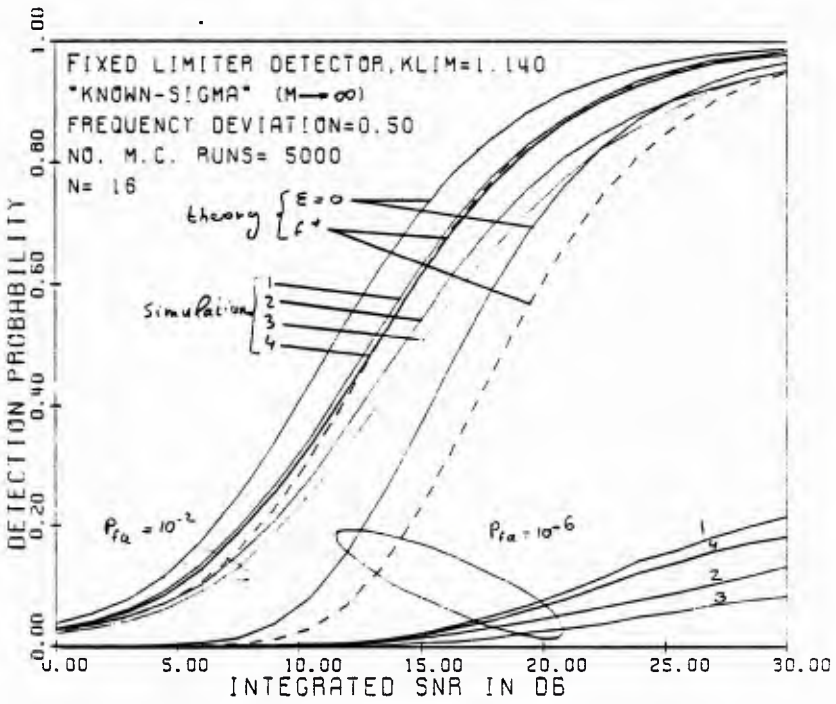


Fig. 5.27 Detection probability. Test - D5, K=1.14,Δf =.5



6. ROBUST RANK TESTS FOR NARROWBAND FADING SIGNALS

6.1 Introduction

6.1.1 - Motivation

On initial reflection, it seems somewhat contradictory to investigate robust rank tests. Since rank tests are generally distribution-free (DF) under the null hypothesis, it is reasonable to ask if their detection performance are not already inherently insensitive to the distribution shape. It turns out that this is indeed the case, but we have not found any previous work pointing to that conclusion. Moreover, it will be shown that by a slight increase in complexity, we can also have maximin optimal performance. Other reasons that motivate us to study tests based on rank statistics are the following.

i) It is possible to design rank tests that attain the DF-CFAR property for all n, under relatively weak assumptions. In particular, some of the tests that will be studied in this chapter only require that the observations be i.i.d. under the null hypothesis.[1] By way of contrast, the various robust SSQME tests of Chapters 4-5 are only asymptotically CFAR. Though the simulation results, Section 5.4, showed very good CFAR property for $n = 16$, this is not true for all n, for all desired P_{fa} level, and for all possible noise contaminations. For instance, in order to achieve $P_{fa} < 10^{-4}$ with contaminations of $\epsilon \approx 0.5$, it was found that trimming levels of $\alpha \rightarrow 0.5$ are required, i.e., the detector is based on median estimators. This leads to large variances as indicated by the asymptotic formula, Eq. (5.3). The variances are even larger for small n; hence, P_d is further degraded.

ii) With the SW-SSQME tests, Section 5.2a, the adaptive threshold estimator, $W(\mathbf{u}, \mathbf{v})$ of Eq. (5.5), does not efficiently utilize the total number of available noise reference samples, nM. Actually, the ensuing detectability loss is a function of M and not of

[1] This is probably the most important feature of target detection systems (radar, sonar), where initiation of some action process based on false-alarm might bear disastrous results.

nM (see Eq. (5.6)), and is quite large for small M. In a non-homogeneous environment (e.g., clutter "edges" and multiple targets, cf. [20]), M must be kept small in order not to violate the assumption that the reference samples are identically distributed as the test cell samples.[2] On the other hand, the two-sample rank tests that we study use the reference samples more efficiently. Furthermore, it is even possible, with proper choice of the modulating signal, to obtain the optimal performance of the locally most powerful (LMP) detector when the noise p.d.f. is known, and to design for a maximin robust rank detector (in terms of P_d) over a mixture class.

iii) With the scale invariant 1S-SSQME tests which are not based on reference samples, Section 5.2b, it was not possible to obtain exact CFAR and maximin P_d though they were numerically shown to be robust over wide parametric families. Moreover, although these tests suffer a rather small detectability loss at the least favorable p.d.f., the simulation study has shown a slow convergence to asymptotic theory. (Intuitively, the reason for that is the necessity to estimate *two* parameters from the data set).

The rank tests which are proposed and studied in this chapter, essentially avoid the above mentioned problems. On the negative side, some of the tests necessitate increased storage and computational loads. Also, the various SSQME tests are all based on shift invariant estimators, and hence once the convergence to the asymptotic theory has been practically reached under H_0, the detection performance under H_1 follows immediately from the asymptotic results, for *all* signal levels. This is clearly exhibited by the simulation results, Section 5.4. By way of contrast, the proposed rank statistics are not shift invariant, and our asymptotic derivations are valid for weak signals in the *nonsingular detection* sense: $na^2 \to c$. While most of the rank statistics that we consider could be replaced by shift invariant *rank estimators,* see [3], without affecting the efficacy

mance, this will again necessitate

[2] The adaptive threshold $W(\mathbf{u}, \mathbf{v})$ could be robustified against interfering signals and clutter edges in a manner similar to that of our previous work [20], and similar improvements can be anticipated.

performance, this will again necessitate robustly estimating their variance. Thus, we will at most obtain the performance of the SSQME tests, but with increased complexity.

6.1.2 - Preview

In Section 6.2, the available observations structure for typical radar/sonar applications is presented, and some rank tests are classified according to the ranking structure. *One-sample* tests do not require noise-reference observations, while *two-sample* tests are based on ranking the entire data-set: the observations obtained from the test cell under question, plus these which originate from a noise reference. In *single-sweep* tests, each observation ("sweep") from the test cell, is ranked with respect to its contemporary noise reference observations, and the final test statistics is a function of the single sweep ranks, accumulated over the illuminating pulse train period. The various rank tests thus represent different levels of performance vs. complexity. Previous relevant results on rank statistics are also summarized.

In Section 6.3, various asymptotically optimal tests for detection of fading narrowband bandpass signals are derived, when the noise p.d.f. is known only in *shape,* i.e., any shift and scale transformations do not affect their performance. The model is given by the p.d.f. under H_1

$$f(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^{n} f(x_i - as_i A\cos\theta)\, f(y_i - as_i A\sin\theta) \prod_{i=n+1}^{n(M+1)} f(x_i)\, f(y_i) \quad (6.1.1)$$

where A is the *random* signal amplitude, normalized such that a^2 represents the unknown input SNR. $\{s_i\}$ is a known signal sequence, and θ is the random phase. The samples are gathered such that the test cell samples are in the first n places. M represents the number of noise-reference vectors, so there are nM noise alone observations. The $\mathbf{x}$ and $\mathbf{y}$ samples are ranked separately, such that R_{Ni}^{x} is the rank of x_i among $\{x_i\}_{i=1}^{N}$, $N = n(M+1)$, and likewise R_{Ni}^{y} , (for a one-sample problem, $M = 0$). The class of tests that we study is of the form

$$\left[\sum_{i=1}^{N} c_i\, a_N(R_{Ni}^{x})\right]^2 + \left[\sum_{i=1}^{N} c_i\, a_N(R_{Ni}^{y})\right]^2 \underset{H_0}{\overset{H_1}{\gtrless}} t_N \quad (6.1.2)$$

It is shown that, for arbitrary distribution of A, this quadrature structure asymptotically maximizes P_d for the model (6.1.1); that for optimality $c_i = s_i$; and there exists an optimal score function $a_N(i, f)$, related to f. In fact, it is the same score function which was found by Hájek [61] for optimal detection of a lowpass deterministic signal. It should be emphasized that the class of tests (6.1.2) is DF-CFAR with the weak assumption that $\{x_i, y_i\}$ are i.i.d. under H_0. No symmetry assumption on f is necessary, and our test does not utilize the sign of the observations, as has been common in previous works, cf. [32], [40].

In communication and target detection systems, the known part of the signal sequence, $\{s_i\}$ is usually at the control of the system designer. The most interesting result of this section is that if a zero mean $\{s_i\}$ is chosen, the performance of an optimal *one-sample* test, is identical with that of the parametric LMP for *completely* known f, and no reference samples are needed. Furthermore, this is simply achieved with phase shifting by π half of the transmitted pulses, in an arbitrary order. The optimal statistics for each of the channels in (6.1.2) then take the form

$$T(\mathbf{x}) = \sum_{\phi_i=0} a_n(R_{ni}^x) - \sum_{\phi_i=\pi} a_n(R_{ni}^x) \tag{6.1.3}$$

and likewise for the $\mathbf{y}$ observations. For single-sweep ranking, the optimal test is of the form

$$\left[\sum_{i=1}^{n} s_i\, a_{M+1}(R^x_{(M+1)i})\right]^2 + \left[\sum_{i=1}^{n} s_i\, a_{M+1}(R^y_{(M+1)i})\right]^2 \underset{H_0}{\overset{H_1}{\gtrless}} t_n \tag{6.1.4}$$

where here $R^x_{(M+1)i}=1,..,M$ is the rank of x_i in $\{x_i, x_{n+i}, x_{2n+i}, \ldots, x_{Mn+i}\}$. The large saving in ranking complexity is compromised for reduced detection performance, since here only if in addition $M \to \infty$ the performance of the optimal one-sample test is achieved. The section concludes with various numerical comparisons of the detectors. In particular, the superiority over the SW-SSQME test (Chapter 5) when the number M of reference vectors is small, is demonstrated.

Traditionally, rank tests have been designed to achieve a DF-CFAR property, and then the detection probability was analyzed at one point in the noise distribution space, usually at the Gaussian, cf. [21], [32], [40], [52-55]. In Section 6.4, we study the detection performance of the various proposed tests, assuming the uncertainty in the noise p.d.f. is modeled by a mixture family. Common score functions are analyzed - median, Wilcoxon and normal scores. It is found that the quadrature rank tests with these score functions are inherently robust from a detectability point of view. In fact, in the worst case, P_d degrades only as some function of $(1-\epsilon)$, whereas heavily tailed contaminations hardly influence the performance. This is in sharp contrast with the parametric UMP scale invariant detector for fading narrowband signal in Gaussian noise, Chapter 2.2, where $P_d \rightarrow 0$ even for small ϵ when the variance of the contaminating p.d.f. is very large.

In Section 6.5, maximin robust rank tests for the model (6.1.1) are found, when the noise p.d.f. belongs to a mixture family. The structure is identical with that of the optimal tests of Section 6.3, but now the score function $a_N(i, f_0)$ is derived from the least favorable distribution f_0, i.e., the one that minimizes Fisher's information. In a similar manner to the SSQME tests, a soft limiter is utilized, but here it operates on the nominal score function of the ranks, and not on the original observations. The proof of maximin optimality parallels that for rank estimators [3], as the performance of both is proportional to the same functional. As usual, the efficiency at the nominal distribution is sacrificed somewhat in order to achieve overall good behavior. No increase in implementation complexity over the traditional rank tests is incurred by the maximin robust test, as it is required only to store a different vector a_N in the look-up table.

Section 6.6 extends the previous tests to handle an unknown signal frequency, a common problem which is always associated with detection of coherent signals, originating from targets of unknown velocity. For the one- and two-sample ranking, we propose a structure similar to that of Chapter 5.3 - first n Doppler channels are generated by

the usual "butterfly" transformation, and then a robust rank test is performed in each of the Doppler channels. It is shown that the asymptotic SNR loss, for signals whose frequency is not matched to either of the discrete down-conversion frequencies, is identical with that of the DFT detector. For the single-sweep structure, an even more promising result is found. It is possible to achieve roughly the same asymptotic performance of the one- and two-sample tests, when the order of frequency conversion and ranking is interchanged! First, the test cell observations are ranked with respect to their adjacent noise reference observations, and then an FFT is performed on the complex vector of the scores of the ranks. Thus, roughly, the complexity is that of the FFT algorithm, $O(n \log n)$, compared to $O(n^2 \log n)$ of the one-sample and SSQME Doppler detectors. The major disadvantage is that fairly large M is required for small losses, thus the single-sweep rank Doppler detector is less appropriate than the one-sample, in a nonhomogeneous environment.

Finally, Section 6.7 presents some of the important results of an extensive Monte-Carlo study. While convergence to the asymptotic theory is slower, in comparison to the SSQME tests, it has been found that, roughly for $n > 32$, most of the conclusions of previous sections are valid. In general, for small number of reference noise vectors M, the robust rank detectors achieve a better P_d than the robust SSQME tests.

As a consequence of the theoretical and simulation results, our tentative conclusion is that the proposed detectors emerge as more than adequate for successfully treating most of the uncertainties that are encountered in the *real-world* radar/sonar detection environment. Furthermore, the current developments in VLSI and VHSIC technologies indicate that implementation in typical systems can be considered practical.

6.2 Definitions and Preliminaries

6.2.1 - The Available Observations and Classification of Rank Tests

Denote by $\{x_{io}\}_{i=1}^{n}$ the samples of the received waveform which contain the information about the useful signal if it is present, and by $\{x_{ij}\}_{i=1}^{n}$, $j = 1,...,M$, the samples from a noise reference channel. Later, we shall focus on narrow-band processes, where x represents both the in-phase and quadrature components.

As a typical example of radar and active sonar systems, the transmitted waveform is an n-long high frequency pulse train, where the pulse repetition interval (PRI) is T, and nT is the total dwell time of the antenna beam at a given spatial direction. The received waveform is passed through a filter which is matched to a single pulse in the train, and the effective pulse width at its *output* is τ. The $\{x_{ij}\}$ are obtained by sampling the matched filter output every τ seconds and, if the possible target location is at a delay of t_0 from the receiver, we have

$$x_{i0} = x(t_0 + (i-1)T)$$

$$x_{ij} = \begin{cases} x(t_0 - (M/2 - j + 1)\tau + (i-1)T) & , \quad j = 1,..., M/2 \\ x(t_0 + (j - M/2)\tau + (i-1)T) & , \quad j = M/2 + 1,..., M \end{cases} \tag{6.2.1}$$

In this way, the M length window in the range coordinate, which is spanned by the resolution cells adjacent to the possible target location, supplies the noise reference channel. The detection process is basically an $L = T/\tau$ hypothesis problem, where the k^{th} hypothesis (presence of a target at $t_0 = k\tau$) is checked sequentially by sliding the above window in the range coordinate.[3]

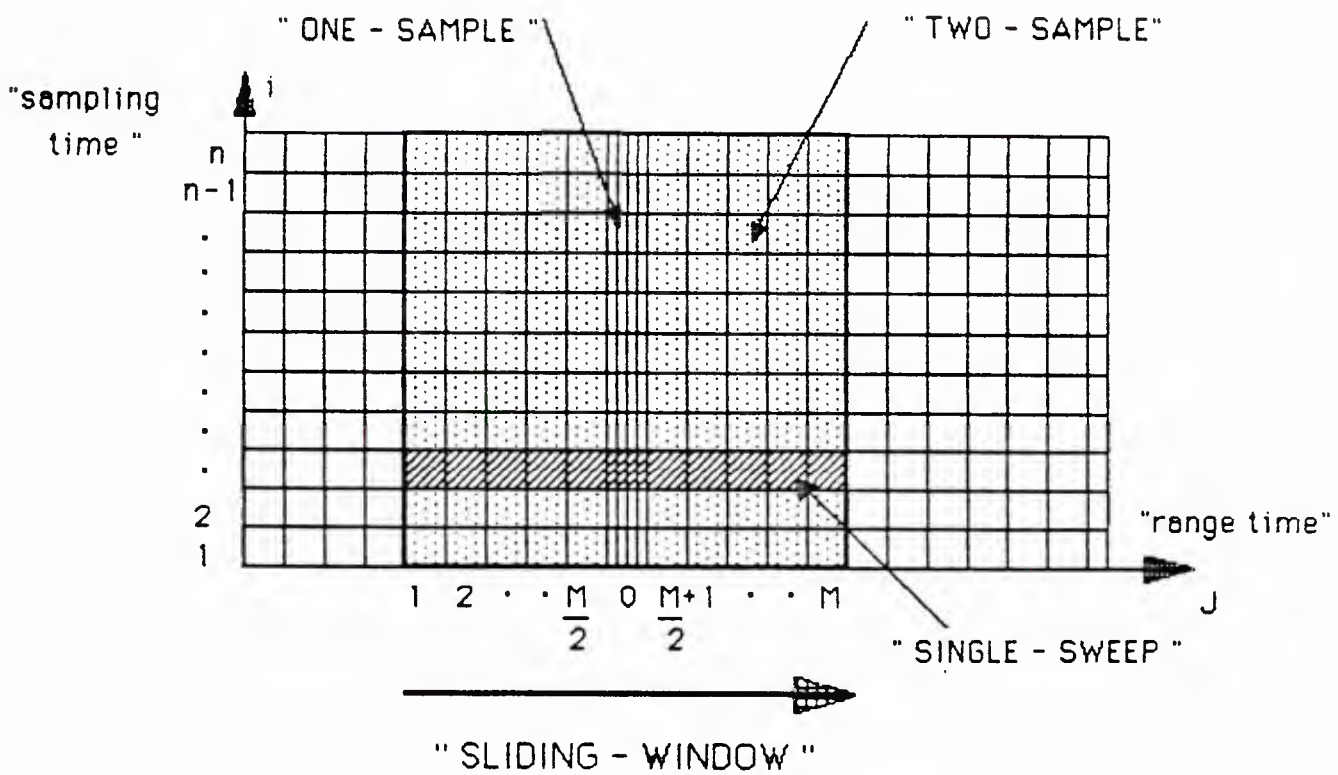
In this chapter, we shall always assume that the observables are i.i.d. across the i index (sampling time) for a given j index ("range time"). On the other hand, the assumption of identical distribution across the j index is less valid in some situations,

[3] Sampling in other coordinates (frequency, spatial direction) is also possible, cf. [18], and will lead to identical results as long as the assumptions about the distribution of the reference samples are identical.

though it might appear strange as $x_{lj}\,(l \neq i)$ is more separated in time from x_{ij} than $x_{ik}\,(k \neq j)$. This is so because the dominant noise sources in radar (sonar) systems are due to clutter (reverberation) reflections of the transmitted waveform from the environment (ground, sea, precipitation clouds, etc.). For a given j, they originate from the same spatial cell, and stationarity can often be assumed as the typical dwell time (O(10 milliseconds)) is too short for substantial change of the reflection characteristics of that cell. On the other hand, samples with different j originate from different spatial cells, and the variation of the reflection characteristics is often large in nonhomogeneous environments. The choice of the reference window length M is an important design parameter. It reflects the designer's compromise between smaller losses at an homogeneous environment with large M, and the losses due to violation of the inherent assumption that the reference window properly represent the noise in the detection cell, c.f. [20].

Various rank test structures are possible, corresponding to the subset of $\{x_{ij}\}$ which is chosen to be ranked, see Figure 6.1.

Figure 6.1



a) Rank Tests for Two-Sample Situations

Sticking to the terminology of the statistical literature, "two-sample" tests are based on ranking of the entire data set. The observables $\{x_{ij}\}$ are pooled in a vector $\mathbf{z}$ of length $N = n(M+1)$, such that the suspected signal samples are in the first group, i.e., $z_{kn+i} = x_{ik}$, $k = 0,1,...M$, $\mathbf{z}$ is then ordered and the vector of ranks $\mathbf{r}$ is generated. Let $z_{(i)}$ be the i^{th} sample in the ordered vector, and r_i the rank of z_i, i.e.,

$$z_i = z_{(r_i)} \ , \ i = 1,2,...n(M+1) \triangleq N \tag{6.2.2}$$

Naive computation of r_i can be done with $O(N^2)$ operations according to

$$r_i = \sum_{j=1}^{N} U(z_i - z_j) \ , \ i = 1,...,N \tag{6.2.3}$$

where $U(\cdot)$ is the unity step function. In practice, random sorting algorithms like the QUICKSORT [39] are preferred, as they require $O(N \log N)$ operations. Throughout, we assume that the p.d.f. of x_{ij} is continuous; thus, the r_i are well defined and unique. The statistic $R_{Ni} = r_i(\mathbf{z})$ will be called the rank of z_i, and $\mathbf{R}_N$ will denote the vector of ranks $(R_{N1}, \ldots, R_{NN})$. A *simple linear rank statistic* for the two-sample situation is defined as

$$T_N(\mathbf{R}_N) = \sum_{i=1}^{n} c_i \, a_N(R_{Ni}) \tag{6.2.4}$$

where $\mathbf{c}$ and $\mathbf{a}_N$ are some N dimensional vectors. Notice that once the data has been ranked, computation of T_N is simpler than that required for the LMP statistics of the form $\sum l(x_i)$, as it is only necessary to store in a look up table the N values of $\mathbf{a}_N$.

Under the alternative, when the SNR is sufficiently large, the ranks of the signal samples $\{x_{i0}\}$ will tend to be the largest, thus T_N will assume larger values than it has under the null hypothesis. We shall see that $\mathbf{c}$ and $\mathbf{a}_N$ can be chosen to optimize the detection process. In general, for a constrained signal structure, two-sample rank tests will result in the best performance, but are also the most demanding in terms of

implementation complexity. Furthermore, when the inherent assumption of identical distribution of the reference samples and the signal samples is not valid, the test will not be CFAR. Also, large detectability losses can occur if, for instance, the noise power in the reference samples is higher than in the signal samples.

Rank tests of this type which are appropriate for radar applications have been studied by Hansen and Olsen [53], Zeoli and Fong [54], and by Al-Hussaini, Badran and Turner [55]; see also [21] and [56] for extensive bibliographies. These works were limited for special non-optimal nonparametric tests (basically, modifications of the Wilcoxon, Mann-Whitney and Savage statistics), and for noncoherent radar processing. The reason for that is, probably, that with coherent signals, a "bank of Doppler tests," as was discussed in Section 5.3, is necessary when the signal frequency is unknown. Therefore, the complexity of coherent rank tests increases by a factor of n, when compared to noncoherent processing. However, as a widely applicable rule, the efficacy of coherent processing is proportional to na^2 (where a^2 is the input SNR), whereas it is proportional to na^4 with noncoherent processing, hence substantially inferior for small SNR (see Appendix H). Also, the above mentioned works have been mostly concerned with the performance for Gaussian derived noises, for which the studied test statistics can be shown to be non-optimal.

b) Rank Tests for the One-Sample Situation

In this case, only the "signal" observables are available, $\{x_{i0}\}_{i=1}^{n}$. In radar-sonar systems, this restriction can be enforced when the environment is known to be nonstationary, and no reliable assumptions concerning the statistical properties of the samples across the reference window can be made. Another useful application of one-sample tests is in a multiple target situation, where a presence of another large signal (e.g., the result of airplanes flying in formation), can cause havoc when a two-sample test is used, cf. [20]. Traditionally, one-sample nonparametric tests have been based on the rank vector $\mathbf{R}_n^+$ of the magnitude of the observations: $R_{ni}^+ = rank$ of $|x_{i0}|$ in

($|x_{10}|,\ldots,|x_{n0}|$), and on the vector of signs

$$T_n(\mathbf{R}_n^+, \mathbf{S}_n) = \sum_{i=1}^{n} sign(x_{i0}) a_n(R_{ni}^+) \tag{6.2.5}$$

To achieve the DF-CFAR property under the hypothesis, it *must* be assumed that the $\{x_{i0}\}$ have symmetric distribution. Application for narrowband coherent signals was studied by Hansen [40] with the Wilcoxon scores $a_n(R_{ni}) = R_{ni}$. In this chapter, we also study statistics which do not depend on the sign vector

$$T_n(\mathbf{R}_n) = \sum_{i=1}^{n} c_i\, a_n(R_{ni}) \tag{6.2.6}$$

Here, no assumption of symmetry is required to obtain CFAR; hence these tests are inherently more robust than the rank-signed tests, (6.2.5). However, in order to achieve satisfactory detection performance, it is necessary to apply a particular modulation of the transmitted signal set $\{s_i\}$, in a rather nonrestrictive way. Perhaps surprisingly, it turns out that it is possible to achieve better performance than that of the two-sample tests, and actually the performance of the LMP parametric detector is reached asymptotically. At the same time, the ranking complexity is substantially reduced. With QUICKSORT, the reduction is by a factor of $(M+1)[1+\log_2(M+1)/\log_2(n)] \approx 9$ for $n=32$, $M=5$.

c) "Single-Sweep" Rank Tests

Here, for each information-gathering "sweep" i, the observables $\{x_{ik}\}_{k=0}^{M}$ are ranked separately, i.e., $R_{(M+1)i}$ = rank of x_{i0} in $\{x_{ik}\}_{k=0}^{M}$; see Figure 6.1. The test statistic is

$$T_N(R_{(M+1)1}, \ldots, R_{(M+1)n}) = \sum_{i=1}^{n} c_i\, a_{M+1}(R_{(M+1)i}) \tag{6.2.7}$$

This test can be considered as a special case of a two-sample problem for randomized blocks [50], or, in accordance with the terminology of Feustel and Davisson [57-58]

as a "mixed" test. Application to nonparametric radar detection, with noncoherent processing and Wilcoxon, Savage and rank-squared scores, have been studied in [52-55]. These tests minimize both the ranking complexity and the storage requirements. From sweep to sweep, it is necessary to keep only a single number, $a(R_i)$, compared to storage of n and $n(M+1)$ observables with the one- and two-samples tests, respectively, before the test statistic can be computed. [4] Further, the ranking complexity is linear in M, since only a single rank is needed for each sweep. For noncoherent processing in Gaussian derived noise, the detectability losses compared to the corresponding two-sample tests are typically negligible if $M > 8$, see [53-55]. This is probably the reason for the wide applicability of these tests in modern noncoherent radar processing.

For a non-stationary environment, the usefulness of these tests is restricted, as M must be kept small, due to the same reasons that were discussed with regard to the two-sample tests.

We will show later that for coherent narrowband signals, the relative loss between the single-sweep and the optimal one-sample test is considerable, when M is restricted to be small in order to preserve the stationarity assumption.

6.2.2 - Some Previous Results on Rank Statistics

In this subsection, we quote and summarize some known properties of rank statistics which are needed in the sequel. Except for the asymptotic distribution results, on which most of the results of subsequent sections are based, the proof of the quoted facts is pretty straightforward. For a complete exposition, see Hájek and Šidák [49].

i) Hypothesis H_0

Under H_0, the observations $\{x_i\}_{i=1}^N$ are assumed to be i.i.d. according to some

[4] The total storage for the entire detection range should be multiplied by the number of range cells $L = T/\tau$, see the discussion following Eq. (6.2.1). L is between 100-1000 for typical radar systems.

arbitrary one-dimensional p.d.f. $f(\cdot)$. Let $\mathbf{R}_N$ be the rank vector of $\mathbf{X}$ and $\mathbf{r}$ some point in the space $\mathbf{P}(N)$ of all permutations of (1,2,...,N). Then

Theorem 6.2.1

$$Pr\{\mathbf{R}_N = \mathbf{r}\} = \frac{1}{N!} \quad , \quad \mathbf{r} \in \mathbf{P}(N) \tag{6.2.8}$$

$$Pr\{R_{Ni} = j\} = \frac{1}{N} \quad , \quad 1 \leq j \leq N \tag{6.2.9}$$

$$Pr\{R_{Ni} = k \, , \, R_{Nj} = l\} = \frac{1}{N(N-1)} \quad , \quad i \neq j \, , \, k \neq l \tag{6.2.10}$$

and so on for the other joint distributions. Notice that these are independent of the particular $f(\cdot)$, thus

Corollary 6.2.1 Any test based on the statistic $T_N \triangleq \sum_1^N c_i \, a_N(R_{Ni})$, is distribution free under H_0 (DF-CFAR).

Define

$$\overline{a} = (1/N) \sum_{i=1}^{N} a_N(i) \, , \, \overline{c} = (1/N) \sum_{i=1}^{N} c_i \, , \, \sigma_a^2 = (1/(N-1)) \sum_{i=1}^{N} (a(i) - \overline{a})^2$$

Then

Theorem 6.2.2 Under H_0,

$$ET_N = N\overline{a} \cdot \overline{c} \quad , \quad VarT_N = \sigma_a^2 \sum_{i=1}^{N} (c_i - \overline{c})^2 \tag{6.2.11}$$

ii) Hypothesis $\mathbf{H}_1$ (Symmetry)

$f(x)$ is an arbitrary one-dimensional symmetric p.d.f., $f(x) = f(-x)$, $x \in (-\infty, \infty)$, and the samples $\{x_i\}_{i=1}^N$ are i.i.d. Let $\mathbf{R}_N^+$ be the rank vector of the absolute values of the observations, $\mathbf{S}$ be the vector of the signs of the observations, and $\mathbf{D}$ be a vector such that $D_i = \pm 1$. Then,

Theorem 6.2.3 Under H_1, for arbitrary symmetric $f(\cdot)$, the random vectors $\mathbf{R}_N^+$ and $\mathbf{S}$ are mutually independent and

$$Pr\{\mathbf{S} = \mathbf{D}\} = (\frac{1}{2})^N \quad , \quad P_r\{\mathbf{R}_N^+ = \mathbf{r}\} = \frac{1}{N!} \quad , \quad \mathbf{r} \in \mathbf{P}(N) \tag{6.2.12}$$

Corollary 6.2.2 Any test based on the statistic $T_N^+ \triangleq \sum_1^N a_N(R_{Ni}^+)\, sign\,(X_i)$, is DF-CFAR under H_1.

Theorem 6.2.4 Under H_1,

$$ET_N^+ = 0 \ , \ \ Var \ \ T_N^+ = \sum_{i=1}^{N} a^2(i) \tag{6.2.13}$$

iii) The ϕ-Functions and Scores

Let $F(x)$ be the distribution of f , and $F^{-1}(u)$ its inverse. Define

$$\phi(u\ ,\ f\ ) = -\frac{f\ '(F^{-1}(u))}{f\ (F^{-1}(u))} \qquad 0 < u\ < 1 \tag{6.2.14}$$

Lemma 6.2.5 If $f\ (\cdot)$ is strongly unimodal (i.e., $-\log f\ (x)$ is convex on the support of $F(x)$), $\phi(u\ ,\ f\ )$ is non-decreasing.

Lemma 6.2.6 If Fisher's information $I(f\ ) < \infty$,

$$I(f\ ) = \int_0^1 \phi^2(u\ ,\ f\ )du \tag{6.2.15}$$

Define the H_0 scores for a density $f\ (\cdot)$ by

$$a_N(i\ ,f\ ) \triangleq E_{H_0}\{-\frac{f\ '(x_{(i)})}{f\ (x_{(i)})}\} = N\binom{N-1}{i-1}\int_{-\infty}^{\infty} f\ '(x)F^{i-1}(x)[1-F(x)]^{N-i}\ dx \tag{6.2.16}$$

where $\{x_{(i)}\}$ are the order statistics of $\mathbf{x}$.

In a similar manner for H_1,

$$\phi^+(u\ ,f\ ) \triangleq \phi(\frac{u+1}{2},\ f\ ) \tag{6.2.17}$$

$$a_N^+(i\ ,f\ ) \triangleq E_{H_1}\{-\frac{f\ '(\mid x\ \mid_{(i)})}{f\ (\mid x\ \mid_{(i)})}\} \tag{6.2.18}$$

Theorem 6.2.7 If $\int \mid f\ '(x)\mid dx\ < \infty$, then

$$\sum_{i=1}^{N} a_N(i\ ,f\ ) = 0 \ , \ \frac{1}{N}\sum_{i=1}^{N} a_N^2(i\ ,f\ ) \leq I(f\ ) \tag{6.2.19}$$

and similarly for a_N^+. As $N \to \infty$, the left-hand side of the second equation converges to $I(f)$.

We shall refer to the $a_N(i,f)$ as the optimal scores (for a *known* f). Let $J(u)$ be some arbitrary function, bounded and square integrable on [0,1]. *Approximate scores* are generated according to

$$a_N(i) = J(i/(N+1)) \tag{6.2.20}$$

or

$$a_N(i) = N \int_{(i-1)/N}^{i/N} J(u)du \tag{6.2.21}$$

The $a_N(\cdot)$ are said to converge quadratically to $J(u)$, denoted as $a_N(\cdot) \xrightarrow{q} J$ if

$$\lim_{N \to \infty} \int_0^1 [a_N(1 + [uN]) - J(u)]^2 du = 0 \tag{6.2.22}$$

where $[x]$ is the largest integer in $[0,x]$.

Theorem 6.2.8 If $a_N(i)$ satisfy either (6.2.20) or (6.2.21), $a_N(\cdot) \xrightarrow{q} J$. Also, $a_N(i,f) \xrightarrow{q} \phi(u,f)$. The implication of this theorem is that, asymptotically, $a_N(i,f)$ can be replaced by the approximate scores $a_N(i)$, with $J(u) = \phi(u,f)$, see [49, pp. 155-166].

iv) Asymptotic Normality of Linear Rank Statistics

The asymptotic distribution of simple linear rank statistics is by no means a straightforward consequence of the central limit theorem - notice that the components of the rank vector are neither independent nor identically distributed. The question of asymptotic normality of rank statistics has been treated by many theoretical statisticians, cf. Chernoff and Savage [59] and the other references cited there. The following deep theorems by Hájek [49,61], who drew on an earlier work of LeCam [60], are the most general for our purposes.

Consider a shift alternative to H_0,

$$f_1(\mathbf{x}) = \prod_{i=1}^{N} f(x_i - d_i) \quad , \quad I(f) < \infty, \tag{6.2.23}$$

where the $\{d_i\}$ satisfy Noether's conditions

$$\lim_{N \to \infty} \sum_{i=1}^{N} (d_i - \overline{d})^2 = b^2 < \infty \tag{6.2.24}$$

$$\lim_{N \to \infty} \frac{\sum_{i=1}^{N} (d_i - \overline{d})^2}{\max_{1 \le i \le N} (d_i - \overline{d})^2} = \infty \tag{6.2.25}$$

Notice that these conditions constitute a special definition of a "small signal," which is an example of the "contiguity" (by LeCam's notion) of $f_1(\mathbf{x})$ to $f_0(\mathbf{x})$. In particular, for either a zero mean signal $d_i = (-1)^i d$, or for a constant signal in the "signal" samples $d_i = d(1-U(i-n-1))$, (6.2.25) is automatically satisfied. The condition (6.2.24) reduces to $Nd^2 \to b$ or $n(1-n/N)d^2 \to b$. Thus, the regularity condition for non-singular detection is sufficient, and $d \to 0$ alone is not necessary.

We are interested in the distribution of

$$T_N(\mathbf{R}_N, \mathbf{c}) = \sum_{i=1}^{N} c_i a_N(R_{Ni}) - N\overline{a} \cdot \overline{c} \tag{6.2.26}$$

where $a_N(\cdot) \xrightarrow{q} J$.

Theorem 6.2.9 Under the alternative (6.2.23), when the conditions (6.2.24) and (6.2.25) hold, the rank statistic (6.2.26), with

$$\lim_{N \to \infty} \frac{\sum_{1}^{N} (c_i - \overline{c})^2}{\max_{1 \le i \le N} (c_i - \overline{c})^2} = \infty \tag{6.2.27}$$

is asymptotically $N(\mu_{dc}, \sigma_c^2)$, where

$$\mu_{dc} \triangleq [\sum_{1}^{N} (c_i - \overline{c})(d_i - \overline{d})] \int_0^1 [J(u)\phi(u, f)du \tag{6.2.28}$$

$$\sigma_c^2 \triangleq [\sum_1^N (c_i - \overline{c})^2] \int_0^1 [J(u) - \overline{J}]^2 du \ , \ \overline{J} \triangleq \int_0^1 J(u) du \qquad (6.2.29)$$

and $\phi(u, f)$ is given by (6.2.14). The same is true under H_0 where $\mu_{dc} = 0$.

For H_1, with alternatives $\prod_{i=1}^n f(x_i - d)$ (recall that now f is symmetric), and $nd^2 \to b^2 < \infty$, similar result hold for the distribution of rank signed statistics.

Theorem 6.2.10 The statistic (6.2.5) with $a_n(\cdot) \xrightarrow{q} J$, is asymptotically $N(\mu_d, \sigma^2)$, where

$$\mu_d \triangleq dn \int_0^1 J(u) \phi^+(u, f) du \qquad (6.2.30)$$

$$\sigma^2 \triangleq \sum_1^n a_n^2(i) \to n \int_0^1 J^2(u) du \qquad (6.2.31)$$

where $\phi^+(u, f)$ is given by (6.2.17).

6.3 Asymptotically Optimal Quadrature Rank Tests for Narrowband Fading Signals in Known Noise

6.3.1 - Two-Sample, One-Sample and Single-Sweep Optimal Tests

Asymptotically optimal rank tests for detection of additive *deterministic* signals have been studied by Hoeffding [62], Capon [63], and Hájek [61]. The results of [61] actually show that the rank information is asymptotically sufficient, i.e., the optimal rank tests achieve the best possible performance. As in Chapter 4, for narrowband fading signals with random parameters, we optimize the performance within a given family of tests. In view of our previous results and Appendix E, it is not surprising that the optimal ensuing test is composed of the test statistics that are also optimal for deterministic signals, and has the structure of a quadrature test.

The joint p.d.f. under the alternative, for the observations structure of Section (6.2.1.a), is

$$f(\mathbf{x},\mathbf{y}) = \prod_{i=1}^{n} f(x_i - a s_i A\cos\theta)\, f(y_i - a s_i A\sin\theta) \prod_{i=n+1}^{N} f(x_i) f(y_i) \qquad (6.3.1)$$

where A is a positive r.v. with arbitrary distribution, θ is $U[0,2\pi]$, $\{s_i\}$ is a *known* signal sequence with

$$\overline{s} \triangleq \frac{1}{n}\sum_{i=1}^{n} s_i < \infty \quad , \quad \overline{s^2} \triangleq \frac{1}{n}\sum_{i=1}^{n} s_i^2 = 1 \qquad (6.3.2)$$

The parameter a corresponds to the unknown SNR and $N = n(M+1)$. $f(\cdot)$ is a p.d.f. of *known shape,* i.e., any shift and scale transformation $f(x) \rightarrow (1/\sigma) f((x-\mu)/\sigma)$ is allowed, as the rank vector is invariant for it.

The following tests will be CFAR for *any* $f(\cdot)$ [see corollaries 6.2.1 and 6.2.2] and will optimize the detection performance for the particular $f(\cdot)$.

Define a class of tests

$$Q_N(\mathbf{R}_N^x, \mathbf{R}_N^y; \mathbf{c}, \mathbf{a}_N) = Q_N(T_N(\mathbf{R}_N^x; \mathbf{c}, \mathbf{a}_N), T_N(\mathbf{R}_N^y; \mathbf{c}, \mathbf{a}_N)) \qquad (6.3.3)$$

where $\mathbf{R}_N^x$ is the vector of ranks for the pooled observations $\{x_i\}_{i=1}^N$, $\mathbf{R}_N^y$ is likewise for $\{y_i\}_{i=1}^N$, $a_N(\cdot) \xrightarrow{q} J(u)$, $J(u)$ is quadratic integrable on $[0,1]$, $\mathbf{c}$ is some vector that satisfies (6.2.27), and $T_N(\cdot, \mathbf{c}, \mathbf{a}_N)$ is a linear rank statistic given by Eq. (6.2.26).

Proposition 6.3.1 Under the alternative (6.3.1) for H_0, when only the shape of f is known, if $\lim_{n\to\infty} a^2 n = b < \infty$, the asymptotically optimum test in the class (6.3.3) is, for any distribution of A,

$$[\sum_{i=1}^{n} s_i\, a_N(R_{Ni}^x, f)]^2 + [\sum_{i=1}^{n} s_i\, a_N(R_{Ni}^y, f)]^2 \underset{H_0}{\overset{H_1}{\gtrless}} t_n(\alpha) \qquad (6.3.4)$$

where $a_N(i, f) \xrightarrow{q} J(u) = \phi(u, f)$ according to either (6.2.16), (6.2.20) or (6.2.21) and

$$t_n(\alpha) = -2\log\alpha\; n\left(1 - \frac{\overline{s}^2}{M+1}\right) \int_0^1 J^2(u)\, du \qquad (6.3.5)$$

In particular, the optimum detection probability for a Rayleigh signal with $\overline{A^2} = 1$ is

$$\beta(\alpha, f, \mathbf{s}) = \alpha^{1/[1 + SNR(f, s)]} \tag{6.3.6}$$

where the "effective integrated SNR" is defined as[5]

$$SNR(f, \mathbf{s}) = \frac{na^2}{2} \left[1 - \frac{\overline{s}^2}{(M+1)}\right] I(f) \tag{6.3.7}$$

Proof We first need to show asymptotic normality of the in-phase and quadrature rank statistics. Condition (6.2.24), specialized to the model (6.3.1) with $\overline{s^2}=1$, reduces to

$$\left(1-\frac{\overline{s}^2}{M+1}\right) \lim_{n\to\infty} na^2 < \infty$$

Also, Eq.(6.2.25) becomes

$$\lim_{n\to\infty} \frac{n\left(1-\overline{s}^2/(M+1)\right)}{\max_i\{(s_i-\overline{s}/(M+1))^2,\ \overline{s}^2/(M+1)^2\}} = \infty$$

Thus, the conditions of Theorem 6.2.9 are satisfied, $T_N(\mathbf{R}_N^x;\ \mathbf{c}, \mathbf{a}_N)$ and $T_N(\mathbf{R}_N^y;\ \mathbf{c}, \mathbf{a}_N)$ are, conditioned on A and θ, jointly Gaussian and independent. Substituting the quadrature shift parameters of (6.3.1) into (6.2.28) yields

$$E\{T_N(\mathbf{R}_N^x;\ \mathbf{c}, \mathbf{a}_N) \mid A, \theta\} = aA\cos\theta \sum_{i=1}^{N} (c_i - \overline{c})(d_i - \overline{d}) \int_0^1 J(u)\phi(u, f)du \tag{6.3.8}$$

$$E\{T_N(\mathbf{R}_N^y;\ \mathbf{c}, \mathbf{a}_N) \mid A, \theta\} = aA\sin\theta \sum_{i=1}^{N} (c_i - \overline{c})(d_i - \overline{d}) \int_0^1 J(u)\phi(u, f)du \tag{6.3.9}$$

where $d_i = s_i(1-U(i-n-1))$, $\overline{d} = (n/N)\overline{s}$, and $V_1\{T_N\} \to V_0\{T_N\}$ are identical for both channels and are given by (6.2.29). As in the proof of Proposition 4.1 (see Chapter 4.2, Eq. (4.16) and the following), the UMP test statistic in the class (6.3.3), for a given $J(u)$ and $\mathbf{c}$, is given by $T_N^2(\mathbf{R}_N^x;\ \mathbf{c},\ \mathbf{a}_N) + T_N^2(\mathbf{R}_N^y;\ \mathbf{c},\ \mathbf{a}_N)$. This is true for *any* distribution of A since the likelihood ratio is a monotone function of this statistic for any positive A. It remains to optimize $\mathbf{c}$ and $J(\cdot)$. As in Chapter 4.2 (see also Appendix E), the

[5] Notice that $SNR(f, \mathbf{s})/(na^2/2)$ coincides with the usual definition of the efficacy for asymptotically Gaussian test statistics. See also Appendix E.

asymptotic detection probability is

$$E_A\, Q\left[\left[\frac{E^2\{T_N(\mathbf{R}_N^x;\,\mathbf{c},\mathbf{a}_N)\mid A,\theta\}+E^2\{T_N(\mathbf{R}_N^y;\,\mathbf{c},\mathbf{a}_N)\mid A,\theta\}}{2V_0\{T_N\}}\right]^{1/2},\sqrt{-2\log\alpha}\right] \tag{6.3.10}$$

where $Q(\cdot,\cdot)$ is Marcum's Q-function (notice that θ cancels). Since it is monotonic increasing in the first argument, it suffices to maximize the efficacy

$$\epsilon(\mathbf{c},\mathbf{d},J(\cdot),f)=G(\mathbf{c},\mathbf{d})G(J(\cdot),f(\cdot))=\frac{[\sum_{i=1}^{N}(c_i-\overline{c})(d_i-\overline{d})]^2}{n\sum_{i=1}^{N}(c_i-\overline{c})^2}\cdot\frac{[\int_0^1 J(u)\phi(u,f)du]^2}{\int_0^1[J(u)-\overline{J}]^2du} \tag{6.3.11}$$

For future purposes, $G(J,f)$ is also given by

$$G(J(\cdot),f(\cdot))=\frac{[\int_{-\infty}^{\infty}J(F(x))f'(x)dx]^2}{\int_0^1[J(u)-\overline{J}]^2du}=\frac{[\int_{-\infty}^{\infty}J'(F(x))f^2(x)dx]^2}{\int_0^1[J(u)-\overline{J}]^2du} \tag{6.3.12}$$

where the second equality is valid for bounded and differentiable $J(\cdot)$.

From the Cauchy-Schwartz inequality, we must have for a maximum

$$c_i-\overline{c}=d_i-\overline{d}=-\frac{\overline{s}}{(M+1)}+s_i(1-U(i-n-1)) \tag{6.3.13}$$

Since $\int_0^1\phi(u,f)=\int_{-\infty}^{\infty}f'(x)dx=0$, we can replace in the numerator of (6.3.11) $J(u)$ by $J(u)-\overline{J}$, and from the Cauchy-Schwartz inequality we must have for a maximum $J(u)=\phi(u,f)$. Hence the optimal scores are $a_N(i,f)\xrightarrow{q}\phi(u,f)$. Substituting (6.3.13) into (6.2.26),

$$T_N=\sum_1^N(c_i-\overline{c})a_N(R_{Ni},f)=\sum_{i=1}^{n}s_i\,a_N(R_{Ni},f)-\frac{\overline{s}}{M+1}\sum_{i=1}^{N}a_N(R_{Ni},f) \tag{6.3.14}$$

and the last term vanishes by virtue of (6.2.19). Also,

$$\sum_{i=1}^{N} (d_i - \overline{d})^2 = \sum_{i=1}^{n} (s_i - \frac{\overline{s}}{M+1})^2 + \frac{Mn}{(M+1)^2}\,\overline{s}^2 = n(1 - \frac{\overline{s}^2}{M+1}) \qquad (6.3.15)$$

Under H_0, Q_N is one sided exponentially distributed, and (6.3.5) is obtained from (6.2.29) and (6.3.15). For a Rayleigh signal, Q_N is again one-sided exponential, as in Chapter 4.3. Combining (6.2.15), (6.3.11) and (6.3.15), yields (6.3.6) and (6.3.7), Q.E.D.

Remarks 1) Proposition 6.3.1 remains valid when the definition of the class (6.3.3) allows different $\mathbf{a}_N$ and $\mathbf{c}$ vectors for the $\mathbf{x}$ and $\mathbf{y}$ channels. 2) Note that in the optimal test, the known signal sequence is correlated with the vector of scores, and *not* with the raw observation vector. This is also true for all the other tests that follow.

Corollary 6.3.2 For a constant signal, $s_i = 1$, $SNR(f, \mathbf{s}) = \frac{na^2}{2}\,\frac{M}{M+1}\,I(f)$. As $M \to \infty$, it approaches the performance of the *parametric* locally most powerful (LMP) quadrature test

$$Q_{LMP} = \left[\sum_{i=1}^{n} s_i \frac{f'(x_i)}{f(x_i)}\right]^2 + \left[\sum_{i=1}^{n} s_i \frac{f'(y_i)}{f(y_i)}\right]^2 \qquad (6.3.16)$$

which necessitates *complete* knowledge of $f(\cdot)$ to obtain the CFAR property. While the rate of convergence with increasing width of the reference window is low, the losses for small M are reasonably low, e.g. 3dB and 1dB for $M = 1$ and 4, respectively.

Proof Computation of the asymptotic detection probability for the test based on the statistic (6.3.16) is similar to that in the proof of Prop. 6.3.1. Substituting $\overline{s}=1$ in (6.3.7) shows they are identical when $M \to \infty$.

By the way we have defined the parameters in (6.3.1), aA represents the unknown reflection characteristics of the target signals, whereas the $\{s_i\}$ are at the control of the system designer. In particular, it is possible to obtain arbitrary $\{s_i\}$ by introducing phase modulations in the transmitted high-frequency pulse-train. In this way, none of the transmitter power, which determine the achievable detectability, is lost. From

(6.3.7) or (6.3.15), it is clear that even better performance than that of Corollary 6.3.1 is possible.

Corollary 6.3.3 (The optimal signal.) With zero mean signal, $\overline{s} = 0$, the best performance can be achieved even with $M = 0$, i.e. *without any reference window!* In particular, for even n, with $(n/2)$ phase shifts ϕ_i of π on the transmitted pulse train, at arbitrary order, one obtains the the simple statistic

$$T_N = \sum_{\phi_i = 0} a_n(R_{ni}, f) - \sum_{\phi_i = \pi} a_n(R_{ni}, f) \tag{6.3.17}$$

Notice that CFAR is obtained without any assumption of symmetry of $f(\cdot)$, and no use is made of the sign of the observations. Therefore, a quadrature test based on (6.3.17) will be the best for multiple targets and non-stationary environment. The statistic (6.3.17) is also very reasonable for finite n. For large SNR, at both quadrature channels, the $\{R_{ni}, \phi_i = 0\}$ will tend to be among the lower or upper half of the rank vector, depending on the arbitrary target phase θ. As $a_n(i)$ is increasing with i, see Lemma 6.2.5, the absolute value of T_N grows with SNR, and achieves its maximum value when $SNR \to \infty$ and the ranking is perfect (i.e., $R_{ni} > n/2$ or $R_{ni} \leq n/2$, $\forall$ $\phi_i = 0$).

Proposition 6.3.4 Under the alternative (6.3.1) for H_1 and a one-sample situation, if $\lim a^2 n < \infty$, the asymptotically optimal test is, for any distribution of A,

$$[\sum_{i=1}^{n} sign(x_i) a_n(R_{ni}^{x+}, f)]^2 + [\sum_{i=1}^{n} sign(y_i) a_n(R_{ni}^{y+}, f)]^2 \underset{H_0}{\overset{H_1}{>\!<}} t_n(\alpha) \tag{6.3.18}$$

where $a_n(i, f) \xrightarrow{q} J(u) = \phi^+(u, f)$ and

$$t_n(\alpha) = -2\log\alpha \; n \int_0^1 J^2(u)\, du \tag{6.3.19}$$

For a Rayleigh signal, $SNR(f) = \frac{na^2}{2} I(f)$.

Proof Parallels the proof of Proposition 6.3.1, using Theorem 6.2.10.

Remarks 1) The performance is identical with the test of corollary 6.3.3. However, this test is more complicated to implement as the sign vectors are also needed, and is less robust as it is CFAR only for symmetric $f(\cdot)$. 2) With $a_N(i) = i$, this is the Wilcoxon narrowband detector, which has been studied by Helstrom, Carlyle and Hansen [40]. However, no optimum properties of this test were discussed by these researchers.

Similarly, for single-sweep ranking, Chapter 6.2.1c,

Proposition 6.3.5 Under the alternative (6.3.1) for H_0, the single-sweep asymptotically optimum test is, if $\lim na^2 < \infty$, for any distribution of A,

$$[\sum_{i=1}^{n} s_i\, a_{M+1}(R^x_{(M+1)i}, f)]^2 + [\sum_{i=1}^{n} s_i\, a_{M+1}(R^y_{(M+1)i}, f)]^2 \underset{H_0}{\overset{H_1}{>\!<}} t_n(\alpha) \qquad (6.3.20)$$

where $a_{M+1}(i, f)$ are given by Eq. (F.8) of Appendix F, and

$$t_n(\alpha) = -2\log\alpha\; n\frac{1}{M+1}\sum_{l=1}^{M+1} a^2_{M+1}(l, f) \qquad (6.3.21)$$

For a Rayleigh signal,

$$SNR(\mathbf{s}, f) = \frac{na^2}{2}\,\frac{1}{M+1}\sum_{l=1}^{M+1} a^2_{M+1}(l, f) \to \frac{na^2}{2}\, I(f) \qquad (6.3.22)$$

Thus, *with large* M, single-sweep rank tests do not incur any loss and are appropriate for homogeneous environment.

Proof See Appendices E and F. Unlike the previous cases of one and two-sample rank tests, the derivation is based on a simpler asymptotic distribution result, as the channels' statistics in (6.3.20) are the sum of i.i.d. variables.

6.3.2 - Some Optimal Score Functions (See also [49]).

i) For Gaussian f, $\phi(u, f) = \phi^{-1}(u)$. When the scores are computed according to (6.2.20), i.e., $a_N(i) = \phi^{-1}(i/(N+1))$, the various rank tests are of the Van der

Waerden (VDW) type. If the scores are computed from (6.2.16), they are of the Fisher-Yates-Terry-Hoeffding (FYTH) type. Both are asymptotically equivalent.

ii) For logistic p.d.f., $f = e^{-x}(1+e^{-x})^{-2}$, $\phi(u,f) = 2(u - 1/2)$. Computed by (6.2.20), the $a_N(i)$ are the Wilcoxon scores $i/(N+1) - 1/2$.

iii) For double exponential f, $\phi(u,f)$=sign$(u-1/2)$. The corresponding two-sample test is of the median type, $T = \sum_{i=1}^{N} \text{sign}(R_{Ni} - \frac{N+1}{2})$. For one-sample situation, $\phi^+(u,f)$=sign$u$ (see 6.2.17), thus $a_n^+(R_i)$=1 and a narrowband sign test is obtained.

iv) For the "least favorable" p.d.f. from an ϵ-mixture family with a Gaussian nominal, Eq. (4.13), $f'/f\,(F^{-1}(u))$ is constant for $|F^{-1}(u)| \leq k$, and in this interval $F(x) = (1-\epsilon)\,\phi(x) + \epsilon/2$; thus $F^{-1}(u)$ can be solved explicitly and it is obtained (see also [3, pp. 99])

$$\phi(u,f) = \begin{cases} k & , \; u \geq \frac{\epsilon}{2} + (1-\epsilon)\,\phi(k) \\ \phi^{-1}(\frac{u-\epsilon/2}{1-\epsilon}) & , \; \frac{\epsilon}{2}+(1-\epsilon)\phi(-k) \leq u < \frac{\epsilon}{2}+(1-\epsilon)\,\phi(k) \\ -k & , \; u < \frac{\epsilon}{2} + (1-\epsilon)\,\phi(-k) \end{cases} \qquad (6.3.23)$$

The various rank tests designed with this score function possess a maximin property as will be shown in Section 6.5. We will refer to them as the robustified VDW type (RVDW). The robustness results from limiting the influence of the extreme ranks. As $\epsilon \to 1$, the RVDW tests clearly reduce to the median tests.

6.3.3 - Comparison Between Two-Sample and Single-Sweep Tests

In general, the relative performance between the tests that are based on different ranking structures, depend on the underlying noise p.d.f. and on the score function.

With the optimal score function for a known f, the ARE is found from (6.3.7) and

(6.3.22) as the ratio of efficacies,

$$ARE_{2S,SS} = \frac{\epsilon_{2S}}{\epsilon_{SS}} = (1 - \frac{\overline{s}^2}{M+1})\, I(f)\left[\frac{1}{M+1}\sum_{l=1}^{M+1} a_{M+1}^2(l,f)\right]^{-1} \quad (6.3.24)$$

where subscripts SS and $2S$ denote single-sweep and two-sample, respectively.

Assuming that $N = n(M+1)$ is sufficiently large (but finite) such that asymptotic normality has been reached, an approximation for the relative efficiency (RE) is obtained when in the above $I(f)$ is replaced by $I_N(f) \triangleq (1/N)\sum_{1}^{N} a^2(l,f)$. Notice that (6.3.24) with $\overline{s}=0$ gives also the ARE with respect to the previously studied one-sample tests.

Table 6.1 displays $I_N(f)$ with VDW scores. Table 6.2 shows the RE for Gaussian f, $n = 20$ and different window sizes. It is seen that in a non-stationary environment, when small M is preferred, the loss incurred by single-sweep ranking is substantial.

$I_N(f)$ for the "least favorable" $f^*(\epsilon = 0.1)$ with the RVDW ($k = 1.14$) scores, generated from (6.3.23) according to (6.2.20), is shown in Table 6.3. As the rate of convergence to $I(f^*) = .671$ is much faster, the loss from single-sweep ranking is smaller. For $n = 20$, $M = 3$, $RE_{1S,SS}$=1.31 only. (Of course, the absolute efficiency for each of the tests for f^* is smaller by 0.671 compared to the Gaussian).

With the Wilcoxon scores, the ARE of single-sweep ranking with respect to the full ranking schemes takes a very simple form:

Proposition 6.3.6 The ARE of single-sweep ranking with Wilcoxon scores is *independent of* f and is given by

$$ARE_{2S,SS} = [1 - \overline{s}^2/(M+1)]\,[1 + 2/M] \quad (6.3.25)$$

$$ARE_{1S,SS} = 1 + 2/M \quad (6.3.26)$$

The proof is given in Appendix G, where it is shown that for $J(u) = u - 1/2$, and $a(i)=i/(N+1) - 1/2$, $\int_0^1 J(u)\phi(u,f)du$ is proportional to $\sum_{i=1}^{N} a(i)a_N(i,f)$ for all

f . As for the VDW scores, for non-stationary environment when large M cannot be taken, Eqs.(6.3.25-26) show that single-sweep ranking incurs considerable loses; nevertheless, the ARE is clearly bounded by 3 ($\rightarrow$5 dB loss).

6.3.4 - Comparison with the SW-SSQME Test

From Eq. (5.6), the detection probability of a Rayleigh signal with the SW-SSQME CFAR detector, *at the noise p.d.f. against which it is optimized,* is given by

$$\beta_{SW}(f\ ,\alpha) = \left[1 + \frac{\alpha^{-1/M} - 1}{1 + SNR_{SW}}\right]^{-M}, \quad SNR_{SW} = \frac{na^2}{2} I(f\ ) \qquad (6.3.27)$$

The detection probability for the two-sample optimum rank test (2SR) is given by (6.3.6), where the SNR with a constant signal (the worst case) is $SNR_{2SR} = \frac{M}{M+1}\frac{na^2}{2}I(f\ )$. As the limit distributions are *different* for finite M, a single-number measure of performance like the ARE is inappropriate. In Table 6.4, we compare the effective integrated SNR, $ISNR \triangleq \frac{na^2}{2}I(f\ )$, which is required to attain specified α and β, equal for both detectors.

$$ISNR_{SW} = (\alpha^{-1/M} - 1)/(\beta^{-1/M} - 1) - 1 \qquad (6.3.28)$$

$$ISNR_{2SR} = \frac{M+1}{M}(\log\alpha/\log\beta - 1) \qquad (6.3.29)$$

Notice that the performance for the optimal one-sample rank tests is given by the $M \rightarrow \infty$ column.

The performance of the optimal rank tests is seen to be substantially superior unless M is large. Therefore, as was promised to be shown at the beginning of this chapter, they are to be preferred in a nonhomogeneous environment where M must be restricted.

N	$I_N(f)$	N	$I_N(f)$
4	.386	100	.923
6	.497	120	.933
8	.570	140	.940
10	.622	160	.946
12	.661	180	.951
14	.691	200	.955
16	.716	400	.974
32	.822	600	.982
64	.892	1200	.990

Table 6.1 $I_N(f) = \frac{1}{N}\sum_{i=1}^{N}[\phi^{-1}(\frac{i}{(N+1)})]^2$. Gaussian f

M	1	3	5	7	9	11	15	31	63
$RE_{2S,SS}$	1.34	1.76	1.56	1.45	1.38	1.33	1.26	1.16	1.09
$RE_{1S,SS}$	2.67	2.35	1.88	1.66	1.54	1.45	1.35	1.20	1.11

Table 6.2 Relative efficiencies of one-sample and two-sample ranking compared to single-sweep ranking. VDW scores, $\overline{S}=1, n=20$.

N	4	6	10	16	32	64	100	200	∞
$I_N(f^*)$	.508	.582	.608	.632	.652	.661	.665	.668	.671

Table 6.3 $I_N(f^*)$ for the least favorable $f^*(\epsilon=0.1)$. RVDW scores with k=1.14.

M	1	2	4	8	16	32	128	∞
$ISNR_{SW}$ (dB)	19.9	13.2	10.2	8.8	8.2	7.8	7.6	7.5
$ISNR_{2SR}$ (dB)	10.5	9.3	8.5	8.0	7.8	7.6	7.5	7.5

Table 6.4 Effective integrated SNR for SW-SSQME and 2-sample rank tests, both optimized for the same noise p.d.f. a) $P_{fa}=10^{-2}$, $P_d=0.5$.

M	1	2	4	8	16	32	128	∞
$ISNR_{SW}$ (dB)	29.5	22.2	19.0	17.6	20.0	16.6	16.4	16.3
$ISNR_{2SR}$ (dB)	19.3	18.1	17.3	16.8	16.6	16.4	16.3	16.3

Table 6.4*b*) $P_{fa}=10^{-2}$, $P_d=0.9$.

M	1	2	4	8	16	32	128	∞
$ISNR_{SW}$ (dB)	60.0	33.8	22.1	17.0	14.8	13.7	13.0	12.8
$ISNR_{2SR}$ (dB)	15.8	14.5	13.7	13.3	13.0	12.8	12.8	12.8

Table 6.4*c*) $P_{fa}=10^{-6}$, $P_d=0.5$.

M	1	2	4	8	16	32	128	∞
$ISNR_{SW}$ (dB)	69.5	42.7	30.6	25.4	23.2	22.2	21.4	21.1
$ISNR_{2SR}$ (dB)	24.2	22.9	22.1	21.7	21.4	21.3	21.2	21.1

Table 6.4*d*) $P_{fa}=10^{-6}$, $P_d=0.9$.

6.4 The Natural Robustness of Rank Tests

The SW-CFAR version of the parametric UMP detector for narrowband fading signals in Gaussian noise, Eq. (2.8), is extremely non-robust. Its effective SNR, which determines the detection probability according to (6.3.6) or (6.3.10), is inversely proportional to the *total* noise variance. In an unrestricted ϵ-mixture family, $V = (1-\epsilon) + \epsilon\sigma_c^2$ can approach infinity for very spiky noise, hence detectability could be completely lost. The robust tests that were studied in Chapters 4 and 5 are immune to heavy tailed contaminations, but once the robust estimator was designed according to some $k(\epsilon_d)$, if the actual contamination ϵ is larger and $\epsilon \to 1$, the effective SNR $\to 0$. From Eq. (5.3), if *all* the contamination lies outside [-k, k], we obtain

$$V = \frac{1+2(k^2-1)\phi(-k)-2kg(k)+k^2\epsilon/(1-\epsilon)}{(1-\epsilon)[1-2\phi(-k)]^2} \underset{\epsilon\to 1}{\longrightarrow} \infty \tag{6.4.1}$$

However, it turns out that the various rank tests *inherently* possess similar immunity against heavy tailed noise, and in fact, the detectability loss is characterized by some power of $(1-\epsilon)$. This is seen from the functional that influences the effective SNR and through it the detection probability, $G(J,f)$ of Eq.(6.3.12).

Proposition 6.4 Assume that $J(u)$ is monotonic increasing, and denote the denominator of (6.3.12) by $\overline{J^2}$. Then, if the noise p.d.f. belongs to a mixture family $f_\epsilon = (1-\epsilon)f_0 + \epsilon h$, we have:

a) If $J(\cdot)$ is differentiable,

$$G(J,f) > (1-\epsilon)^4[\min_u J'(u)]^2[\int f_0^2(x)dx]^2/\overline{J^2} \tag{6.4.2}$$

b) If $J(u)$ is piecewise constant and discontinuous at $u = 1/2$, with a jump $\Delta J(1/2)$,

$$G(J,f) = \frac{\Delta J^2(1/2) f^2(0)}{\overline{J^2}} > \frac{(1-\epsilon)^2 \Delta J^2(1/2) f_0^2(0)}{\overline{J^2}} \tag{6.4.3}$$

Proof Immediate from the definition and the assumptions, using $f^2 > (1-\epsilon)^2 f_0^2$. For Part a, from the variational necessary condition for a minimum of G with the constraints on f, it can be shown that no solution for it exists and thus the inequality in (6.4.2) can never be replaced by an equality. It can also be shown that as the spread of the contamination increases, G approaches its lower bound, which is reached with degenerate contaminations (e.g., $h = U[-a,a]$, $a \to \infty$ or $N(0,c^2)$, $c \to \infty$).

Applying the proposition for the various score functions, for a Gauss-Gauss mixture $F_\epsilon(x;c) = (1-\epsilon)\phi(x) + \epsilon\phi(x/c)$, we obtain after solving the integrals

i) VDW Scores $\overline{J^2} = I(\phi) = 1$, $\min J'(\cdot) = \sqrt{2\pi}$,

$$G(\phi^{-1}(u), f_\epsilon) = 2\pi \left[\int_{-\infty}^{\infty} \exp\left\{ \frac{1}{2} [\phi^{-1}(F_\epsilon(x;c))]^2 \right\} f_\epsilon^2(x;c)\, dx \right]^2 > \frac{(1-\epsilon)^4}{2} \tag{6.4.4}$$

This bound is not tight; for small ϵ, a Taylor expansion yields

$$G(\phi^{-1}(u), f_\epsilon) = [1 - \frac{\epsilon}{2}(c^2 - 1)]^2 + O(\epsilon^3(c-1)) \tag{6.4.5}$$

ii) Wilcoxon Scores, $\overline{J^2} = 1/12$, $J'(\cdot) = 1$,

$$G(u - 1/2, f_\epsilon) = \frac{3}{\pi} \left[(1-\epsilon)^2 + \frac{2\sqrt{2\epsilon(1-\epsilon)}}{(1+c^2)^{1/2}} + \frac{\epsilon^2}{c} \right]^2 > \frac{3}{\pi}(1-\epsilon)^4 \tag{6.4.6}$$

iii) Median Scores $\overline{J^2} = 1$, $J(0) = 1$, using part b),

$$G(\text{sign}(u - 1/2), f_\epsilon) = 4 f_\epsilon^2(0) = \frac{2}{\pi}[1 - \epsilon + \frac{\epsilon}{c}]^2 > \frac{2}{\pi}(1-\epsilon)^2 \tag{6.4.7}$$

iv) RVDW Scores The robustified VDW scores of Eq. (6.3.23), designed according to $F^*(\epsilon_d)$, where ϵ_d is the design ϵ,

$$G(RVDW, f_\epsilon) = \frac{8\pi}{(1-\epsilon_d)^2 I(F^*(\epsilon_d))} \left[\int_0^{x'} \exp\left\{ \frac{1}{2} \left[\phi^{-1} \left[\frac{F_\epsilon(x;c) - \epsilon_d/2}{1-\epsilon_d} \right] \right]^2 \right\} f_\epsilon^2(x;c)\, dx \right]^2 \tag{6.4.8}$$

$$> \frac{1}{2} \frac{(1-\epsilon)^4}{(1-\epsilon_d)^3} \frac{[1 - 2\phi(-x')]^2}{1 - 2\phi(-k)}$$

where x' is the solution of

$$(1-\epsilon)\phi(x') + \epsilon\phi(x'/c) = (1-\epsilon_d)\phi(k) + \epsilon_d/2 \tag{6.4.9}$$

Generally, for $\epsilon > 1-\sqrt{2/3} = 0.18$ and heavily tailed contaminations, the median tests are the most robust (with the largest loss $2/\pi = -1.96$dB at the Gaussian), the Wilcoxon tests are the second best (with merely $3/\pi = -0.2$dB loss at the Gaussian), and the VDW are the least robust. The RVDW is intermediate to the VDW and the median tests.

More specific conclusions can be drawn from Tables 6.5-6.8, which exhibit the exact losses, compared to the UMP for Gaussian, in dB: $-10\log_{10}(G(J, f_\epsilon))$. ((6.4.4) and (6.4.8) were numerically integrated). Notice that the performance of the RVDW tests is *not* identical with that of the SSQME tests, as the functionals $G(J, f)$ and $1/V(l, f)$ of Eq. (4.11) are different, unless both $J(\cdot)$ and $l(\cdot)$ are the optimal for f. However, the difference is generally small, by comparison of Tables 6.8 and 5.1.

ϵ \ c	1.00	3.33	10.00	33.3	100.00
0.01	0.20	0.30	0.35	0.37	0.37
0.10	0.20	1.25	1.75	1.95	2.00
0.20	0.20	2.33	3.44	3.88	4.01
0.30	0.20	3.45	5.27	6.04	6.28
0.40	0.20	4.59	7.26	8.49	8.87
0.50	0.20	5.74	9.43	11.30	11.91

Table 6.5 losses (dB) for Wilcoxon detector, $f = (1-\epsilon)N(0, 1) + \epsilon N(0, c^2)$

ϵ \ c	1.00	3.33	10.00	33.3	100.00
0.01	0.00	0.22	0.36	0.41	0.43
0.10	0.00	1.67	2.56	2.89	2.99
0.20	0.00	3.01	4.67	5.34	5.53
0.30	0.00	4.26	6.75	7.83	8.16
0.40	0.00	5.43	8.87	10.48	11.00
0.50	0.00	6.53	11.04	13.40	14.21

Table 6.6 losses (dB) for VDW detector, $f = (1-\epsilon)N(0, 1) + \epsilon N(0, c^2)$

ϵ \ c	1.00	3.33	10.00	33.3	100.00
0.01	1.96	2.02	2.04	2.05	2.05
0.10	1.96	2.59	2.78	2.85	2.87
0.20	1.96	3.27	3.68	3.83	3.88
0.30	1.96	4.01	4.69	4.95	5.02
0.40	1.96	4.81	5.84	6.23	6.34
0.50	1.96	5.70	7.15	7.72	7.90

Table 6.7 losses (dB) for median detector, $f = (1-\epsilon)N(0,\ 1)+\epsilon N(0, c^2)$

ϵ \ c	1.00	3.33	10.00	33.3	100.00
0.01	0.06	0.18	0.24	0.27	0.27
0.10	0.06	1.37	2.16	2.48	2.57
0.20	0.06	2.74	4.33	4.96	5.14
0.30	0.06	4.04	6.45	7.47	7.78
0.40	0.06	5.25	8.59	10.15	10.64
0.50	0.06	6.40	10.81	13.09	13.86

Table 6.8 Losses (dB) for RVDW rank detector. $f = (1-\epsilon)N(0,\ 1)+\epsilon N(0,\ c^2)$
a) k= 1.945, design ϵ=0.01, $G(f^*)$=–0.27 *dB*

ϵ \ c	1.00	3.33	10.00	33.3	100.00
0.01	0.44	0.52	0.55	0.56	0.56
0.10	0.44	1.26	1.57	1.68	1.72
0.20	0.44	2.15	2.89	3.19	3.28
0.30	0.44	3.14	4.52	5.15	5.35
0.40	0.44	4.24	6.60	7.86	8.25
0.50	0.44	5.43	9.07	10.95	11.57

Table 6.8*b* k= 1.140 design ϵ=0.1, $G(f^*)$=–1.73 *dB*

ϵ \ c	1.00	3.33	10.00	33.3	100.00
0.01	0.77	0.84	0.86	0.87	0.87
0.10	0.77	1.49	1.72	1.81	1.84
0.20	0.77	2.26	2.81	3.01	3.07
0.30	0.77	3.11	4.07	4.45	4.56
0.40	0.77	4.05	5.57	6.24	6.46
0.50	0.77	5.09	7.45	8.68	9.11

Table 6.8*c* k= 0.862 design ϵ=0.2, $G(f^*)$=–3.11 *dB*

ϵ \ c	1.00	3.33	10.00	33.3	100.00
0.01	1.05	1.12	1.14	1.14	1.15
0.10	1.05	1.72	1.94	2.01	2.03
0.20	1.05	2.45	2.92	3.10	3.15
0.30	1.05	3.24	4.05	4.36	4.45
0.40	1.05	4.11	5.36	5.86	6.01
0.50	1.05	5.07	6.92	7.74	8.00

Table 6.8*d* k= 0.685 design ϵ=0.3, $G(f^*)$=–4.51 *dB*

ϵ \ c	1.00	3.33	10.00	33.3	100.00
0.01	1.29	1.35	1.37	1.38	1.38
0.10	1.29	1.94	2.14	2.21	2.23
0.20	1.29	2.64	3.08	3.24	3.29
0.30	1.29	3.40	4.15	4.43	4.51
0.40	1.29	4.24	5.37	5.81	5.93
0.50	1.29	5.16	6.80	7.47	7.68

Table 6.8*e* k= 0.555 design ϵ=0.4, $G(f^*)$=–6.02 *dB*

ϵ \ c	1.00	3.33	10.00	33.3	100.00
0.01	1.51	1.57	1.59	1.59	1.60
0.10	1.51	2.15	2.34	2.41	2.43
0.20	1.51	2.84	3.26	3.42	3.46
0.30	1.51	3.58	4.30	4.56	4.64
0.40	1.51	4.40	5.47	5.88	6.01
0.50	1.51	5.31	6.84	7.46	7.64

Table 6.8*f* k= 0.436 design ϵ=0.5, $G(f^*)$=–7.73 *dB*

ϵ \ c	1.00	3.33	10.00	33.3	100.00
0.01	1.73	1.80	1.81	1.82	1.82
0.10	1.73	2.37	2.56	2.62	2.64
0.20	1.73	3.05	3.46	3.61	3.66
0.30	1.73	3.79	4.48	4.73	4.80
0.40	1.73	4.60	5.63	6.02	6.13
0.50	1.73	5.49	6.95	7.52	7.71

Table 6.8*g* k= 0.291 design ϵ=0.65, $G(f^*)$=–10.96 *dB*

6.5 Asymptotically Maximin Rank Tests

The effective SNR of the optimal rank tests of Section 6.3 were shown to be proportional to $I(f)$. Therefore, the least favorable p.d.f. of Eq. (4.13) is a candidate for a maximin solution. Notice, however, that $G(J,f)$ of Eq. (6.3.12) is not a convex functional, and the maximin property is therefore not a consequence of the minimax property for M-estimators. Huber treated *rank estimators* for location in [3]. It turns out that their variance is precisely $1/G(J,f)$. Thus the minimax property of the estimation variance is translated to a maximin relation on the detection probability. As the proof in [3] is not detailed, we complete it here.

Proposition 6.5.1 Let $\mathbf{P}_\epsilon = \{f = (1-\epsilon)g + \epsilon h\}$, where g is known, symmetric and strongly unimodal, and h an arbitrary symmetric p.d.f. Denote by $f_0 = (1-\epsilon)g + \epsilon h_0$ the p.d.f. which minimize $I(f)$ in $\mathbf{P}_\epsilon$, Eq. (4.13). Let $a_N(i) \xrightarrow{q} J_0(u) = -f_0'(F_0^{-1}(u))/f_0(F_0^{-1}(u))$. Then,

a) (Huber)
$$G(J_0,f_0) \leq G(J_0,f) \quad \forall f \in \mathbf{P}_\epsilon \tag{6.5.1}$$

b) For arbitrary positive r.v. A in (6.3.1),

$$\beta(f) = E_A E_\theta \beta(f \mid A,\theta) \geq \beta(f_0) \quad \forall f \in \mathbf{P}_\epsilon \tag{6.5.2}$$

i.e., combining with Propositions 6.3.1 and 6.3.4, the various one- and two-sample rank tests are asymptotically maximin robust.

Proof a) Since the denominator of (6.3.12) is not a function of f, it suffices to show

$$\int_0^1 J_0'(u) f_0(F_0^{-1}(u))du \leq \int_0^1 J_0'(u) f(F^{-1}(u))du$$

By the assumptions of symmetry and strong unimodality, $J' \geq 0$ and both integrands are symmetric around $u = 1/2$. Also, $J_0'(u) = 0$ outside $I_u(\epsilon) \triangleq \{u \in [1/2, F_0(k(\epsilon))]\}$, from Eq. (4.13). Thus, it suffices to show $f_0(F_0^{-1}(u)) \leq f(F^{-1}(u))$ on $I_u(\epsilon)$. Since $h_0 = 0$ on this interval,

$f_0(F^{-1}(u)) \leq f(F^{-1}(u))$. But on $I_x(\epsilon) \triangleq \{x \in [0,k(\epsilon)]\}$, $F_0(x) \leq F(x)$, and $F_0(0) = F(0) = 1/2$. Thus, on I_u, $F_0(F^{-1}(u)) \leq F(F^{-1}(u)) = u$, and taking F_0^{-1} on both sides, $0 \leq F^{-1}(u) \leq F_0^{-1}(u) \leq k$.

Finally, by virtue of the strong unimodality, $f_0(x)$ is decreasing on $[0,\infty)$. Combining all these inequalities,

$$f_0(F_0^{-1}(u)) \leq f_0(F^{-1}(u)) \leq f(F^{-1}(u))$$

on $I_u(\epsilon)$, Q.E.D.

b) Immediate from (6.3.10), as $\beta(f \mid A)$ is monotonic increasing in $G(J_0,f)$.

Corollary 6.5.2 The rank tests designed according to the RVDW scores, Eq. (6.3.23), are asymptotically maximin robust within a mixture family with $g = N(0,1)$.

Remarks 1) The maximin is valid only for symmetric f. The two-sample and the one-sample tests with zero mean signal, Eq. (6.3.17), remain DF-CFAR for **all** f. 2) For the single-sweep tests, Proposition 6.5.1 is valid if also $M \to \infty$, since the functional in Eq.(F.10) of App.F converges to $G(J,f)$.

6.6 Robust rank detectors for unknown signal frequency

When the signal frequency is unknown, due to Doppler shift of the reflected signal from a target of unknown velocity in a radar/sonar situation, the situation is similar to that of Chap. 5.3. In this chapter, two different ways of constructing Doppler bank detectors are investigated, corresponding to the different ranking structures. We show that the the effective SNR is given by the same expression as before, multiplied by a loss factor for unmatched frequency. Hence, all previous conclusions about maximin optimality remain valid.

6.6.1 One - and two - sample rank Doppler detectors.

The distribution of the observations is given by Eq.(6.3.1), where $\theta \rightarrow \beta_i$

$$\beta_i = 2\pi i \frac{(f + \Delta f_r / n)}{f_r} + \theta \quad , \quad |\Delta| \leq 0.5 \tag{6.6.1}$$

and f_r is the sampling frequency. As with the SSQME - Doppler test, we take $f = (k/n) f_r$, $k = 1,.., n$, and construct n frequency channels by the "butterfly" transformation

$$I_i(k) = \mathrm{Re}[(x_i + jy_i) W_n^{ik}] \quad , \quad Q_i = \mathrm{Im}[(x_i + jy_i) W_n^{ik}] \tag{6.6.2}$$

where $W_n = \exp(-j\, 2\pi/n)$. Thus, Δ is the deviation of the signal frequency, from the center of the nearest filter, normalized by the filter width f_r/n. Notice the "butterfly" transformation is performed only on the test cell, but not on the noise reference observations. For each of the frequency channels, the transformed observations are ranked according to

$$\begin{aligned} R_{Ni}^{I} &= \text{rank of } I_i \text{ in } \{I_1, \ldots\ldots, I_n, x_{n+1}, \ldots\ldots, x_N\} \\ R_{Ni}^{Q} &= \text{rank of } Q_i \text{ in } \{Q_1, \ldots\ldots, Q_n, y_{n+1}, \ldots\ldots, y_N\} \end{aligned} \tag{6.6.3}$$

where the k (frequency) index is dropped for simplicity of notation. The test statistics for each of the frequency channels are of the form $T = I^2 + Q^2$, where

$$I \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i\, a_N(R_{Ni}^{I}) \quad , \quad Q \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i\, a_N(R_{Ni}^{Q}) \tag{6.6.4}$$

and $\{s_i\}$ are the known signal modulations, Eq.(6.3.2). The asymptotic performance of this detector structure is given by the following.

Proposition 6.6.1 a) In the expression for the efficacy, Eq.(6.3.11), the first term $G(\mathbf{s}, \mathbf{s})$ is replaced by

$$G(\mathbf{s}, \mathbf{s}, \Delta) = \lim_{n \to \infty} \frac{\left| \frac{1}{n} \sum_{i=1}^{n} (s_i - \overline{S}/(M+1)) s_i \exp(j\, 2\pi i\, \Delta/n) \right|^2}{(1 - \overline{S}^2/(M+1))} \tag{6.6.5}$$

b) For a zero mean signal, $\overline{S}=0$,

$$G(\mathbf{s},\,\mathbf{s},\,\Delta) = \lim_{n\to\infty}\left|\frac{1}{n}\sum_{i=1}^{n} s_i{}^2\exp(j\,2\pi i\,\Delta/n\,)\right|^2 \tag{6.6.6}$$

which is identical to the loss factor of the quadrature matched filter detector (DFT), for a time varying signal with unmatched frequency. In particular, when $s_i=\pm 1$

$$G(\mathbf{s},\,\mathbf{s},\,\Delta) = [\sin(\pi\Delta)/\pi\Delta]^2 \tag{6.6.7}$$

c) For two - sample rank Doppler tests, with $s_i=1$,

$$G(\mathbf{s},\,\mathbf{s},\,\Delta) = \frac{M}{M+1}[\sin(\pi\Delta)/\pi\Delta]^2 \tag{6.6.8}$$

The proof of a) is given in Appendix I; b) and c) are straightforward consequences. Note that a stronger statement about the asymptotic P_d would be desired. However, we could not prove or disprove that it is given by substituting the modified efficacy into (6.3.10). This was true if I and Q of (6.6.4) were shown to be asymptotically uncorrelated, see App. I. Notice that I_i is uncorrelated with Q_i, but in general for non - Gaussian $\{X_i,\,Y_i\}$ they are not independent. The difficulty is in finding a manageable expression for $P\{R^I_{Ni}=l\,,\,R^Q_{Nj}=m\}$, which is needed to compute $E(IQ)$. Due to the ranking operation, R^I_{Ni} and R^Q_{Nj} are not i.i.d., even though X_i and Y_j are. However, simulation results indicated the stronger statement is probably valid, besides being valid for Gaussian noise.

6.6.2 Single - sweep rank Doppler detectors

Though a Doppler - bank of detectors could be constructed in the previous manner with identical results, for single - sweep ranking a simpler solution in terms of implementation complexity exists. It turns out that it is possible to interchange the order of ranking operation with the frequency down - conversion, without any degradation in the performance ! Specifically, let

$$R^{x}_{(M+1)i} = \text{rank of } x_{i0} \text{ in } \{x_{i0}, x_{i1}, \ldots\ldots, x_{iM}\}$$
$$R^{y}_{(M+1)i} = \text{rank of } y_{i0} \text{ in } \{y_{i0}, y_{i1}, \ldots\ldots, y_{iM}\} \tag{6.6.9}$$

A complex vector of the scored single - sweep ranks is generated,

$$z_i = u_i + jv_i \quad , \quad u_i = a_{M+1}(R^{x}_{(M+1)i}) \quad , \quad v_i = a_{M+1}(R^{y}_{(M+1)i}) \tag{6.6.10}$$

and the test statistic for the k^{th} frequency channel is taken as the DFT

$$T_n(k) = \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i W_n^{ik} \right|^2 \quad , \quad k = 1, \ldots\ldots, n \tag{6.6.11}$$

Notice that since only a *single rank* at each sweep (the index i) is needed, the ranking complexity is linear - $O(M)$, as the i^{th} rank can be computed from $R^{x}_{(M+1)i} = \sum_{k=0}^{M} U(x_{i0} - x_{ik})$. $\{T_n(k)\}$ can be computed by a FFT algorithm, and with the same simplifying assumptions of Chap. 5.3, the complexity of the FFT part of the detector is $O(5n \log_2 n + 4n)$. Combining, the total complexity is $O(2nM + 5n \log_2 n + 4n)$. This is in sharp contrast with the complexity of the previous structures, which is $O(2n^2 \log_2 n + 10n^2 + 4n)$ for one - sample ranking (with QUICKSORT algorithm), and even more for the two - sample detector. In App. I, we prove the following result on the asymptotic performance:

Proposition 6.6.2 For $na^2 \rightarrow c^2$, the limiting detection probability of the test statistic (6.6.11) is

$$\beta = E_A \, Q\left[cAG^{1/2}(\mathbf{a}_{M+1}, f) \, | \sin \pi\Delta / \pi\Delta | \, , \, \sqrt{-2 \log \alpha} \right] \tag{6.6.12}$$

where $Q[\cdot, \cdot]$ is Marcum's Q - function and

$$G(\mathbf{a}_{M+1}, f) = \frac{\left[\frac{1}{M+1} \sum_{l=1}^{M+1} a_{M+1}(l) a(l, f) \right]^2}{\frac{1}{M+1} \sum_{l=1}^{M+1} a_{M+1}^2(l)} \tag{6.6.13}$$

As a consequence, all previous conclusions on maximin optimality remain valid.

Why does the DFT after ranking scheme work in this situation ? The key factor is the independence of the $\{R_{(M+1)i}\}$, as each observation is ranked only with respect to its adjacent noise reference samples. Thus, when the SNR is large, when signal of frequency k is present, the $\{u_i\}$ and $\{v_i\}$ oscillate, with a period corresponding to k. Of course, all higher harmonics of k are present, due to the nonlinear operation $a(R_i(x_{i0}))$. The essence of the above result is asymptotic linearization, i.e., the first harmonic is the dominant term, and therefore the same loss factor of the DFT detector is obtained. The same can not be true for the other ranking structures, due to the strong dependency in the vector of ranks.

6.7 Finite sample simulation results

6.7.1 The cases studied

The performance under the hypothesis and the alternative was investigated by means of Monte-Carlo (MC) simulations, for various quadrature rank tests, under different noise situations.

One- and two-sample tests have been simulated, with the Wilcoxon(WX), normal scores according to Van der Waerden(VDW) version, and the robustified VDW (RVDW) score functions (see Sections 6.3.2-6.3.4.). In the following, the notation VDW(n,M) will designate a VDW test, with n main samples and nM noise-reference samples. For the RVDW test, a third number in the brackets will indicate the factor k at which the score function is limited, see Eq.(6.3.23). M = 0 will designate a one-sample test with a zero mean known signal sequence, according to Eq.(6.3.17). One-sample tests that are based in addition on the signs of the observations, were not studied. As discussed before, they are inferior under situations of non-symmetric noise p.d.f. At this time, we do not yet have results for single-sweep rank quadrature tests.

For the detection performance, 8 different Gauss-Gauss noise mixtures $(1-\epsilon)N(0,\,1) + \epsilon N(0,\,c^2)$ have been simulated. They will be denoted f_i according to Table 6.7.1. As is evident from the table, all contaminated cases represent large, heavy tailed deviations from the nominal assumption, with the total noise power increased by 10 to 20 dB.

f_i	ϵ_i	c_i	total noise variance in [dB]
f_1	0.0	—	0
f_2	.01	30.04	10
f_3	.1	9.54	10
f_4	.1	31.48	20
f_5	.2	6.78	10
f_6	.2	22.27	20
f_7	.4	4.85	10
f_8	.4	15.76	20

Table 6.7.1 The studied noise p.d.f.'s.

6.7.2 Performance under the null hypothesis

Under the hypothesis, as long as all the observations are i.i.d., the various tests studied are distribution free. Therefore, a major simplification and saving in execution time was possible - instead of generating the original observables and ranking them, the random rank vector was generated directly, in the following manner. An integer in $[1,..., n]$ was drawn with a uniform probability $1/N$, stored as R_1, and deleted from the original group of integers. At the second stage, R_2 was drawn from the reduced group, independently of R_1, with a uniform probability $1/(N-1)$, and so on. For the two-sample tests, only $n << N=n(M+1)$ ranks were generated, as the other ranks are not required for computation of the test statistic. In this way, the probability distribution of the rank vector is as required by (6.2.2). From the vector of ranks, the various quadrature test statistics were generated according to (6.3.4) or (6.3.17), with the previously mentioned score functions. The known signal sequence was $s_i=1$ for two - sample tests, and $s_i=1$ or -1 (equal number), according to (6.3.17) for one - sample tests.

About 45 cpu minutes were required on the IBM/3081, with $N=64$, $2\cdot10^6$ repetitions of the Monte - Carlo experiment, where all 3 detectors (WX, VDW, RVDW) simulated simultaneously on the same vector of ranks. On the DEC-VAX/750, about 35 cpu *hours* were required for the same simulation size [1].

Figures 6.7.1-6.7.3 exhibit $-\log(P_{fa})$ vs. the threshold multiplier t, for tests of the form $T_N \underset{H_0}{\overset{H_1}{\gtrless}} tE_0(T_N)$, where T_N is the quadrature statistic on the left hand side of either Eq.(6.3.4) or (6.3.17). The RVDW curves are with k = 1.14. $E_0(T_N)$ is calculated from (6.2.11). In this way, a convenient universal comparison of P_{fa} for different

[1] A unique variant of the Importance - Sampling technique, suitable for the class of detectors studied here, was also developed. Although in one case it enabled us to obtain reliable thresholds down to $P_{fa}=10^{-6}$ with only 20,000 repetitions, a 100 fold saving, many trial and error recursions were required until a satisfactory distorted probability measure was found to generate smooth estimates of P_{fa}. The distorted measure depended on the sample size and score function, thus large human effort was needed to cover all the desired cases. Since computer resources have been readily available, most of the thresholds were obtained from the simpler "brute - force" simulation.

sample sizes is possible, and convergence to the asymptotic result is measured by the distance of the Monte- Carlo curves from the straight line. Explicitly, the normalizing factor $E_0(T_N)$ is given by:

$2n^2M(N+1)/12$ - WX, two - sample

$2n^2(n+1)/12$ - WX, one - sample

$2\overline{a_N^2}nM/(M+1)$ - VDW and RVDW, two - sample

$2\overline{a_n^2}n$ - VDW and RVDW, one - sample

where $\overline{a_N^2}=(1/N)\sum_{i=1}^{N} a_N^2(i)$. Numerical values of $\overline{a_N^2}$ are given in Tables 6.1 and 6.3 of Section 6.3.3.

The figures show that with all score functions, for ($n=16$, $M=0$) there is a large deviation from the asymptotic prediction, when $P_{fa}<10^{-2}$. This is not surprising, since the rank statistics are bounded r.v.'s. For this case, it can be calculated that $\max\{T_n/E_0(T_n)\}$ is a number between 11 to 13 for the various score functions. Asymptotically, $t(P_{fa}=10^{-6})=13.8$, thus the asymptotic theory can not predict correctly for this low level. The cases (32,2) and (64,0) essentially converge, while (32,0) and (16,2) are in between. With respect to the score function, RVDW converges faster, though the difference is small. By way of comparison, the SW-SSQME robust test with n = 16 essentially converged to the asymptotic result, see Chapter 5.4. This is so because the M- or L- estimators are "closer" to sum of i.i.d. r.v.'s than the rank statistic.

Thresholds for the P_d simulation were taken from these graphs, with some smoothing for $P_{fa}=10^{-6}$. In the P_d simulation, the sensitivity to deviation in t was measured. As result of the smoothing, we estimate the accuracy of the P_d simulation within ± 0.5dB for $P_{fa}=10^{-6}$ (i.e., the true SNR required to achieve given P_d is within this range), and essentially error free for $P_{fa}>10^{-4}$.

Fig.6.7.1 False alarm probability, Wilcoxon quadrature rank tests.

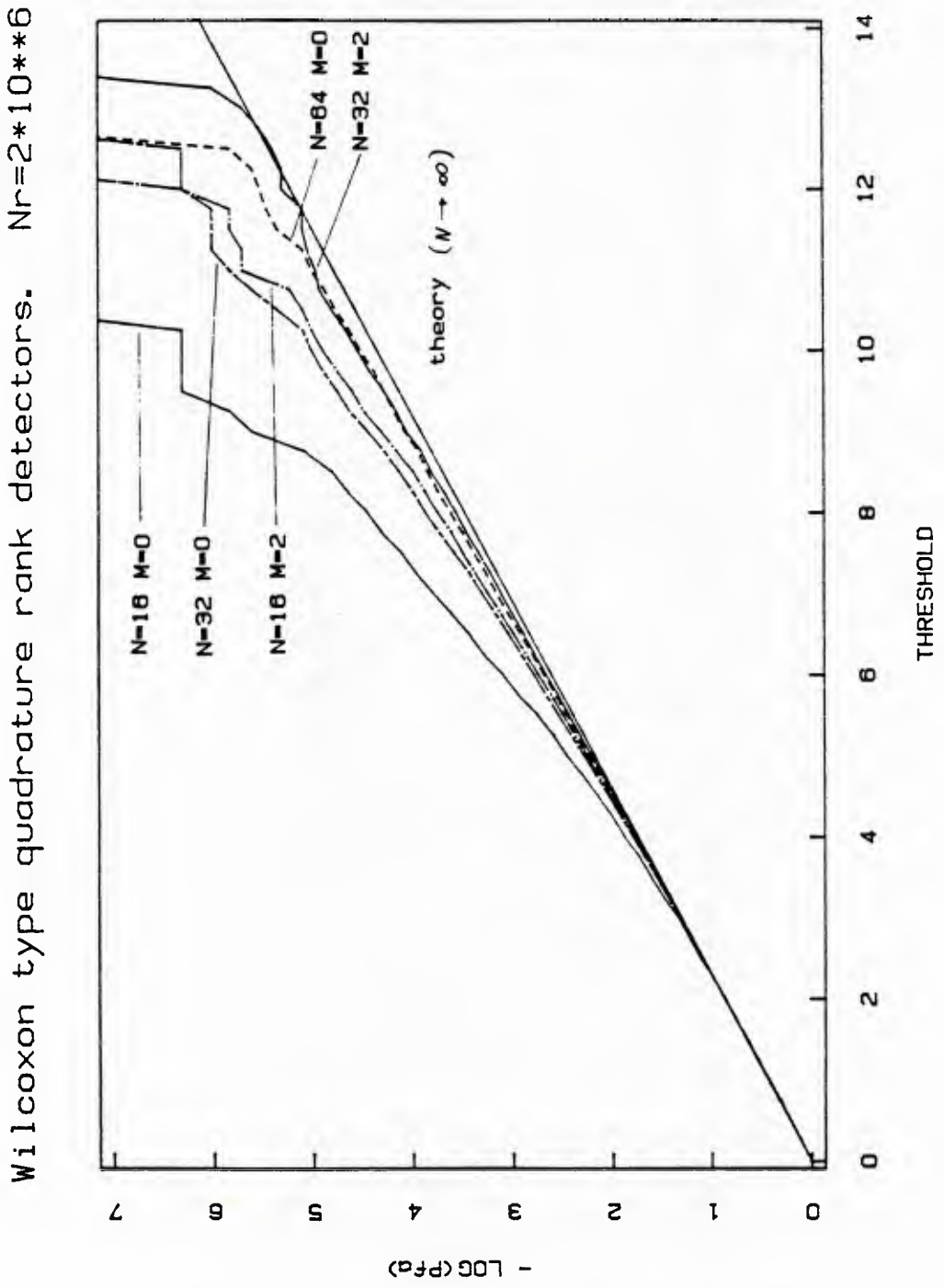


Fig.6.7.2 False alarm probability, VDW quadrature rank tests.

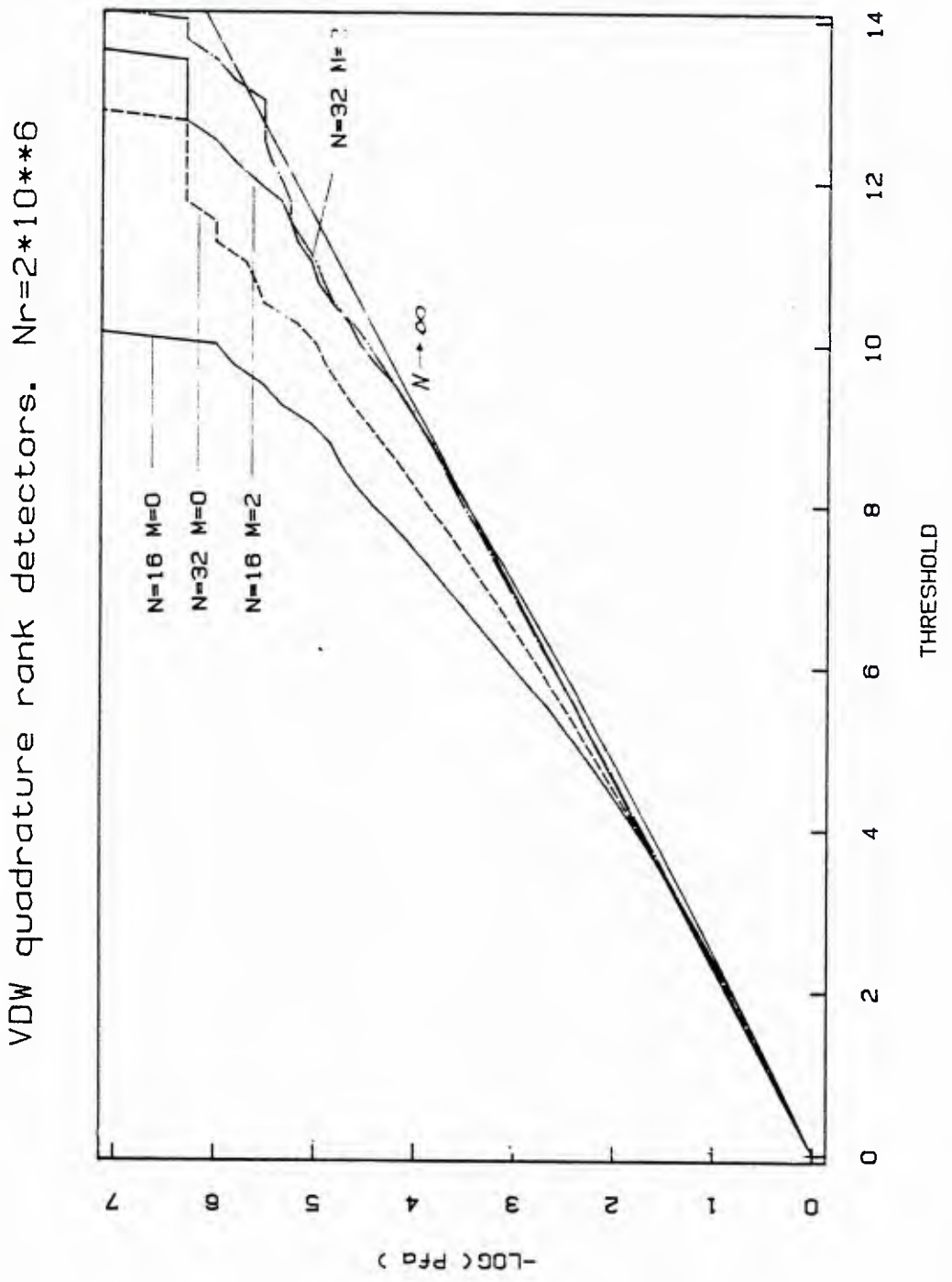
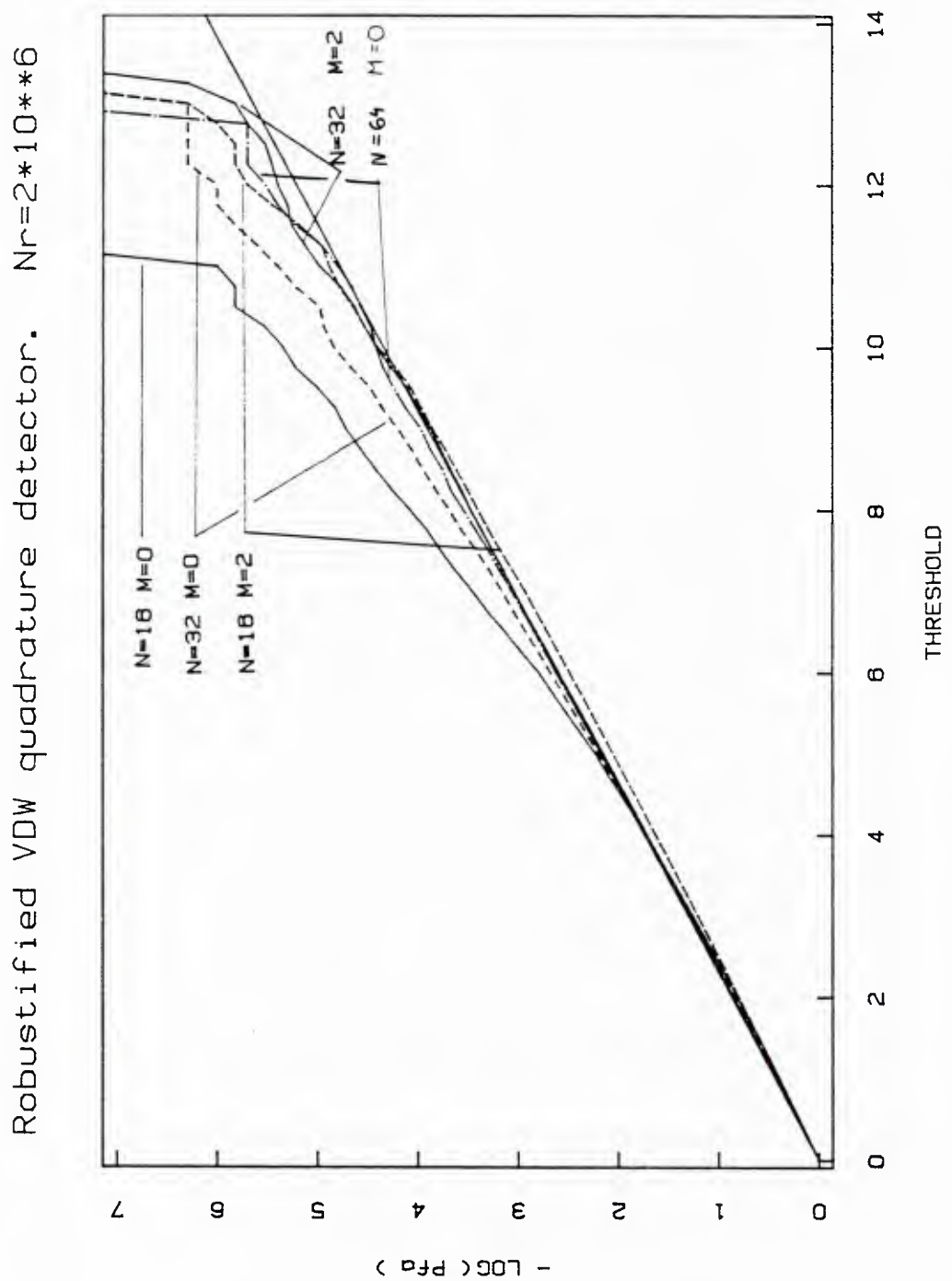


Fig.6.7.3 False alarm probability, RVDW quadrature rank tests.



6.7.3 Probability of detection performance

The simulations for detection probability implement directly the model (6.3.1) and some of the various detectors of Sections 6.3, 6.5 and 6.6. Independent in-phase and quadrature noise r.v.'s with distributions f_1–f_8 are generated by a switching device, as in section 5.4. The in-phase and quadrature components of a Rayleigh signal with uniform phase are added to the noise, two rank vector are generated from the observations, and the test statistics are computed and compared to the previously found thresholds. For most of the results presented here, the signal frequency is assumed known and the "butterfly" transformation, Eq.(6.6.2), is not performed. Figs. 6.7.21-6.7.23 correspond to the Doppler rank detectors of Section 6.6.1.

In a typical run, the 3 different detectors (WX, VDW, RVDW) of equal (n,M) are simulated on the same observations, at 3 P_{fa} levels (10^{-2}, 10^{-4}, 10^{-6}), and at 20 different SNR values. On the IBM/3081, about 4 cpu minutes were required, for n=32 and M=0, with 4 different noise distributions and 2500 Monte-Carlo (MC) repetitions.

In all the following figures, P_d is plotted vs. the nominal integrated SNR $na^2/2$, where $a^2/2$ is the input SNR at the *nominal* component of the mixture. Thus, the actual input SNR is lower by a factor of $1-\epsilon+\epsilon c^2$. This normalization enables us to display all the results on the same range, and it will be seen that this is the dominant factor which determine the detectability, even for heavily tailed contaminations (recall the results of Section 6.4). In all figures, the curves marked (UMP,ϵ=0) correspond to the UMP detector for narrowband slow fading Rayleigh signal in narrowband Gaussian noise of known variance; those marked (theory, $f^*(\epsilon)$) are computed from the asymptotic P_d at the least-favorable p.d.f. If no value of ϵ is indicated, ϵ=0.1

In order to restrict the number of displayed results, we shall concentrate on the RVDW detector. Its performance as function of the sample and reference sizes will be shown, as well as the dependency on the noise distribution. For some cases, comparison with the other detector laws (WX, RVDW) will be given.

Fig. 6.7.4 display P_d curves of RVDW(k=1.14), with various (n,M), against Gaussian noise. In general, good agreement with the asymptotic result will be achieved if the curves are shifted to the right, from the (UMP,ϵ=0) curve, by the dB amount corresponding to the efficacy loss, independent of P_d. It is seen that at $P_{fa}=10^{-2}$, all curves essentially converge in shape; the loss for (64,0) is 0.5-0.8 dB, compared to the efficacy loss of 0.44 dB, see Tbl. 6.8b. Also, the one-sample (16,0) detector is better than the two-sample (16,2) by about 1.5 dB, in agreement with $M/(M+1)=$ 1.8 dB (see Eq. 6.3.29). On the other hand, as expected, deviation from asymptotics is larger at $P_{fa}=10^{-6}$. Here, the distribution of RVDW(16,0,1.14) is very different in shape from the asymptotic result, Eq. (6.3.6). For the other sample sizes, the shape of the asymptotic distribution is virtually obtained, with somewhat increased dB losses. In this case, the (16,2) detector is substantially better than the (16,0) detector (about 9 dB at $P_d=0.8$), contrary to the asymptotic result. The dependency of the RVDW detector on sample size at the f_3 distribution (see Table 6.7.1) is shown in Fig. 6.7.5. Here, agreement with the asymptotics is worse. For the other distributions, all the large ϵ cases (f_3–f_8) behaved similar to Fig. 6.7.5, while f_2 behaved very much like the Gaussian. Similar qualitative behavior was found with the VDW and WX detectors. In fact, our version of the one-sample WX detector, behaves very much like Hansen's narrowband detector [40] which utilizes also the signs of the observations.

Roughly speaking, at $P_{fa}=10^{-6}$, the performance agrees with the asymptotic prediction, with somewhat increased SNR losses, for $n \geq 32$ and small ϵ, for all score functions. For larger contaminations, the breakpoint at $P_{fa}=10^{-6}$ is around $n=32$. At $P_{fa}=10^{-2}$, the (16,0) detectors also come close to the asymptotic prediction, if $\epsilon \leq 0.01$.

The RVDW(16,0,1.14) performance at $P_{fa}=10^{-4}$ against f_1–f_4 noises is shown in Fig. 6.7.6. Notice that the P_d degradation is ranked according to the total noise variance, and for the noises with equal variance (f_2, f_3) P_d is lower when ϵ is larger, as could be expected. We found it possible to improve somewhat the performance against

the larger contaminations by either of the following two modifications. In the first, by taking the RVDW(0.862) score function (it is asymptotically maximin robust when ϵ=0.2) - i.e., by introducing harder limiter on the score function. In the second, we took $a(1)=a(16)=0$ - i.e., the influence of the extreme ranks was totally eliminated. As could be expected, this improvement also resulted in higher losses at the nominal and f_2 distributions. Comparison between the different detector laws for the above parameters at f_4 is shown in Fig. 6.7.7. At $P_d=0.5$, RVDW is better than WX by 1 dB, which in turn is better than VDW by 1.3 dB. These differences are even somewhat larger than the asymptotic ones, compare Table 6.8b with Tables 6.5 and 6.6. Very similar superiority of RVDW was obtained for f_3, with the difference negligible for f_2. Notice that the VDW detector, which is the asymptotically optimal rank detector for Gaussian noise, is quite robust anyway, in agreement with the discussion of Section 6.4. This is in sharp contrast with the total breakdown of the parametric LMP detector under non Gaussian heavy tailed noise.

The performance of RVDW(32,0,1.14) for all P_{fa} levels against f_1–f_4 is shown in Figs. 6.7.8-9. Since convergence to the asymptotics is almost reached at $P_{fa}=10^{-2}$, the curves are almost bounded by the theoretical P_d for the least-favorable distribution. The detection performance is worse at $P_{fa}\leq 10^{-4}$, though the detector is clearly robust. In Fig. 6.7.8 we also demonstrate that the Gaussian noise provides the most difficult detection environment, among noises with equal variance. This is demonstrated by the dashed rightmost curves, which correspond to the UMP detector against narrowband Gaussian noise, with variance equal to that of f_2–f_4.

Fig. 6.7.10 corresponds to the two-sample RVDW(32,2,1.14) detector, where it is seen that convergence to the asymptotics is essentially reached also for $P_{fa}=10^{-6}$. At $P_{fa}=10^{-2}$, the MC losses are 2.4 dB for f_1–f_2, compared to 2.3 and 2.4 dB, respectively, from Table 6.8b (adding the 1.8 db loss appropriate to $M/(M+1)$). For f_3–f_4, the MC loss is 3.4 dB, and is identical with the efficacy loss. Notice that at $P_{fa}=10^{-6}$

the two-sample detector is better than the equivalent one-sample detector, Fig. 6.7.9, which contradicts the asymptotic prediction. A possible explanation is that the two-sample detector, which use more observations, almost converges to the asymptotics at this low level, while the one-sample does not.

Comparison of the different detector laws with (32,0) at f_3 noise is shown in Fig. 6.7.11, and against the f_4 noise in Fig. 6.7.12. Again, the quality ranking is in accordance to asymptotic results, and the WX detector is only slightly inferior to the RVDW. For the large contamination of f_4, RVDW is superior to VDW by about 6 dB at $P_d=0.7$.

Fig. 6.7.13 depicts P_d of RVDW(32,0,.862), which is asymptotically maximin robust for $\epsilon=0.2$, against f_1, f_4–f_6 noises. Recall $\epsilon=0.2$ for f_5–f_6. While the theoretical curve for the least favorable $f^*(0.2)$ does not bound the MC curves at the highly contaminated noises, the performance can be considered satisfactory if the large increase in the input noise variance (10-20 dB) is accounted for. Comparison between different detector laws at $(32,0,10^{-4})$ is shown in Figs. 6.7.14-6.7.16. At the Gaussian, as predicted by Tables 6.5 and 6.6, WX and VDW are essentially identical, and superior by about 0.8 dB over RVDW(0.862), an excellent agreement with Table 6.8c. For f_5–f_6, however, RVDW is better than VDW by about 2.5 and 7 dB, respectively, at $P_d=0.7$. WX's performance is in between.

Fig. 6.7.17 corresponds to RVDW(64,0,1.14) against f_1–f_4. Comparing it with RVDW(32,0,1.14), Fig. 6.7.9, we observe an improvement in convergence at $P_{fa}=10^{-6}$. Now, all P_d curves agree in shape, though the losses in f_3–f_4 are still higher than the efficacy losses.

Fig. 6.7.18 shows P_d of RVDW(64,0,.55), which is maximin robust for $\epsilon=0.4$. The loss at the Gaussian is increased by merely 0.5 dB when compared to the RVDW(64,0,1.14) detector, even less than the efficacy loss of Table 6.8e, but now the robustified detector is seen to handle reasonably well the noises f_7–f_8, with $\epsilon=0.4$. The

large advantage of RVDW over VDW at the heavily tailed noise f_8, about 10 dB at P_d=0.6, is clearly observed in Fig. 6.7.20. Comparison of Fig. 6.7.19 with 6.7.20 also reveals that the heavier tailed noise has a rather limited influence on RVDW, but it causes a higher degradation in WX and VDW, in accordance with the asymptotic results; compare Tables 6.5, 6.6 with 6.8e.

The performance of the one-sample rank Doppler RVDW(32,0,1.14) detector, Section 6.6.1, is exhibited in Figs. 6.7.21-6.7.23. Fig. 6.7.21 corresponds to a matched frequency case, Δ=0. Comparing it with Fig. 6.2.9, we observe slightly lower P_d at the Gaussian and f_2 noises, but some improvement for the heavier tailed f_3–f_4 noises. Similar behavior was found with the WX and VDW detectors. These empirical results support the conjecture following Proposition 6.6.1. The improved performance with higher contaminations can be heuristically explained by the following: I_i and Q_i, which are weighted sums of two r.v.'s, have the same variance as x_i and y_i, but their marginals have shorter tails as consequences of the convolution.

The frequency mismatch cases are shown in Fig. 6.7.22 (Δ=0.25) and Fig. 6.7.23 (Δ=0.5). The latter corresponds to the crossover frequency between adjacent "filters" in the Doppler-bank detector. Notice that the theoretical curves are shifted to the right from those of Fig. 6.7.21 by $(\sin\pi\Delta/\pi\Delta)^2$=0.9dB and 3.9 dB, respectively. Thus, they correspond to the linear DFT detector and to the asymptotic distribution of the rank Doppler RVDW detector (against f^*), under identical frequency mismatch. At $P_{fa}=10^{-2}$, the MC curves are essentially shifted by the appropriate amount from those in Fig. 6.7.21, validating Prop. 6.6.1. At $P_{fa}=10^{-6}$, where convergence to the asymptotics is not yet reached, the losses are somewhat larger than the above values. Nevertheless, as before, the RVDW detector was found superior over the WX and VDW at all heavy tailed noises by roughly the same amount as in previous cases.

6.7.4 Comparison with the robust SSQME detectors

The lack of convergence of the RVDW(16,0) detector to the asymptotic prediction (recall Figs. 6.7.4-6.7.7) might seem to imply that it is inferior to the various robust SSQME detectors. However, a fair comparison must be based on the same number of noise reference vectors. Fig. 6.7.24 shows RVDW(16,4,1.14), and Fig. 6.7.25 corresponds to the SW-SSQME detector with 4 reference vectors and α=0.3 (it is Fig. 5.17 repeated). Notice that the f_4 noises are not the same in the two figures. It is seen that RVDW is superior to SW-SSQME, by about 2 dB at $P_{fa}=10^{-2}$ and 8 dB at $P_{fa}=10^{-6}$. These differences are in good agreement with the asymptotic prediction of Tables 6.4a-d: with M=4, the theoretical SNR difference is 1.7 dB for ($P_{fa}=10^{-2}$, $0.5 \leq P_d \leq 0.9$), 8.4 dB for ($P_{fa}=10^{-6}$, $P_d=0.5$), and 8.7 dB for ($P_{fa}=10^{-6}$, $P_d=0.9$). Recall from Section 6.3.4 that they have different limiting distributions, hence the SNR loss is function of P_d. A detailed comparison of the integrated SNR ($na^2/2$), required for given P_{fa} and P_d, is shown in Table 6.7.2. We thus conclude the RVDW is superior to the SW-SSQME.

P_{fa}	P_d	detector type	f_1	f_2	f_3
10^{-2}	0.5	RVDW(16,4,1.14)	8.9	9.1	10.1
		SW-SSQME(16,4,0.3)	10.8	10.9	11.4
	0.9	RVDW(16,4,1.14)	17.8	17.8	19.3
		SW-SSQME(16,4,0.3)	19.8	19.8	20.2
10^{-6}	0.5	RVDW(16,4,1.14)	14.6	14.8	16.3
		SW-SSQME(16,4,0.3)	22.6	22.8	23.3
	0.9	RVDW(16,4,1.14)	21.1	21.3	25.9
		SW-SSQME(16,4,0.3)	31.2	31.2	31.8

Tbl. 6.7.2 Comparison of the ISNR, $na^2/2$ in dB, between SW-SSQME and RVDW detectors.

Comparison between the SSQME test without reference samples, Fig. 5.20, and the RVDW(16,0,1.14) is less conclusive. At $P_{fa}=10^{-2}$ against f_1–f_2 noises, the performance is roughly the same. For f_3 at this level, the SSQME detector is better, where the SNR loss increases with P_d. At $P_{fa}=10^{-6}$, the P_d curves are totally different. For f_1 and f_2, the curves cross at about $P_d=0.4$, where the SSQME is better for higher values and the difference increases with P_d. At f_3, SSQME is substantially better for all $P_d \geq 0.05$. At $P_{fa}=10^{-4}$, the situation is even more complicated, but in general, for $P_d > 0.9$, the one-sample SSQME is always better and thus should be preferred, at least for the relatively restricted contaminations against which it was simulated ($\epsilon \leq 0.1$).

Fig. 6.7.4 RVDW test, Gaussian noise, convergence with n and M.

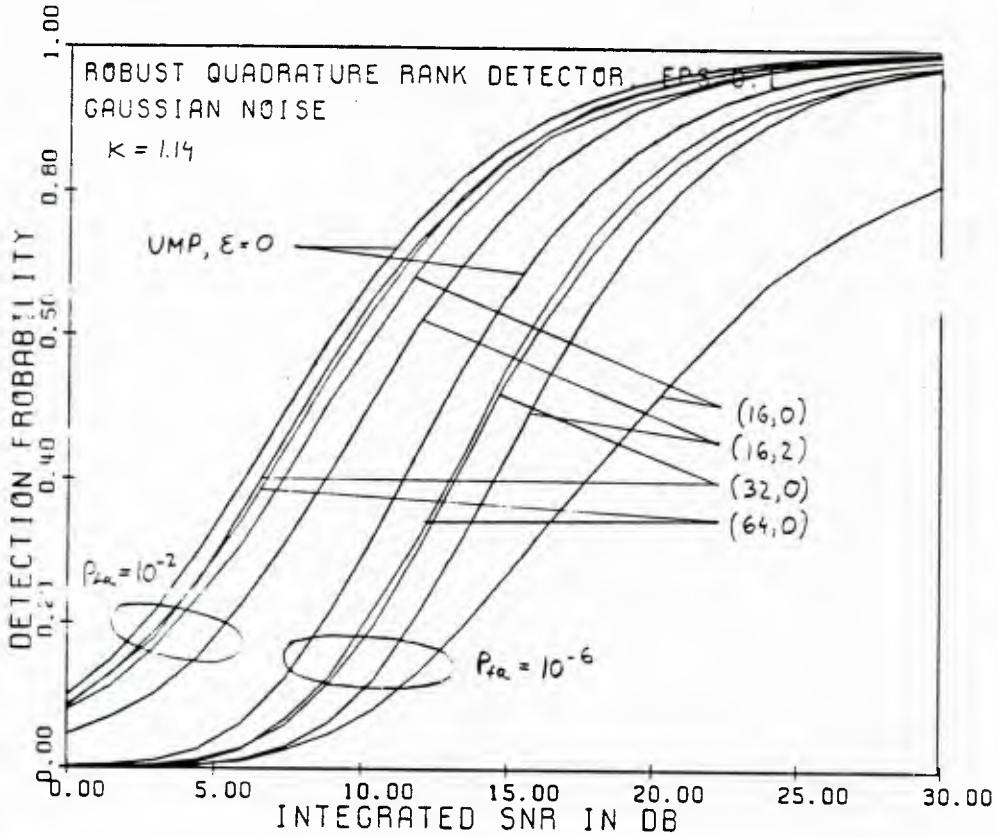


Fig. 6.7.5 RVDW test, f_3 mixture noise, convergence with n and M.

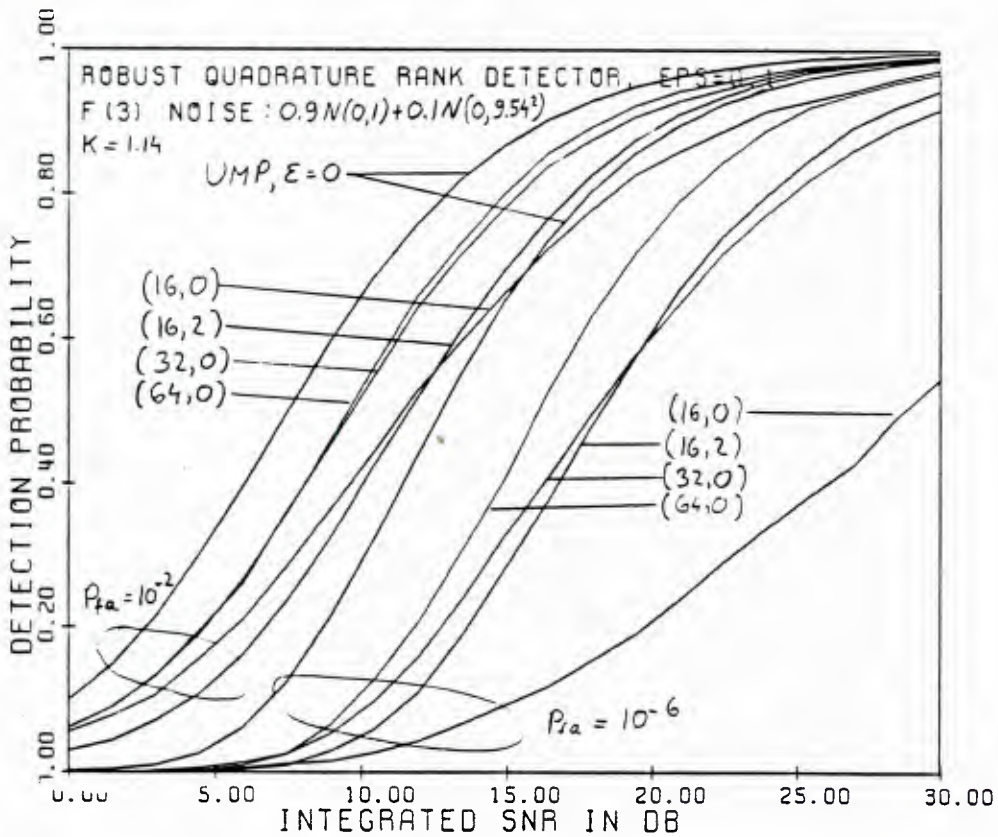


Fig. 6.7.6 1S-RVDW test, f_1 to f_4 noises, $n = 16$.

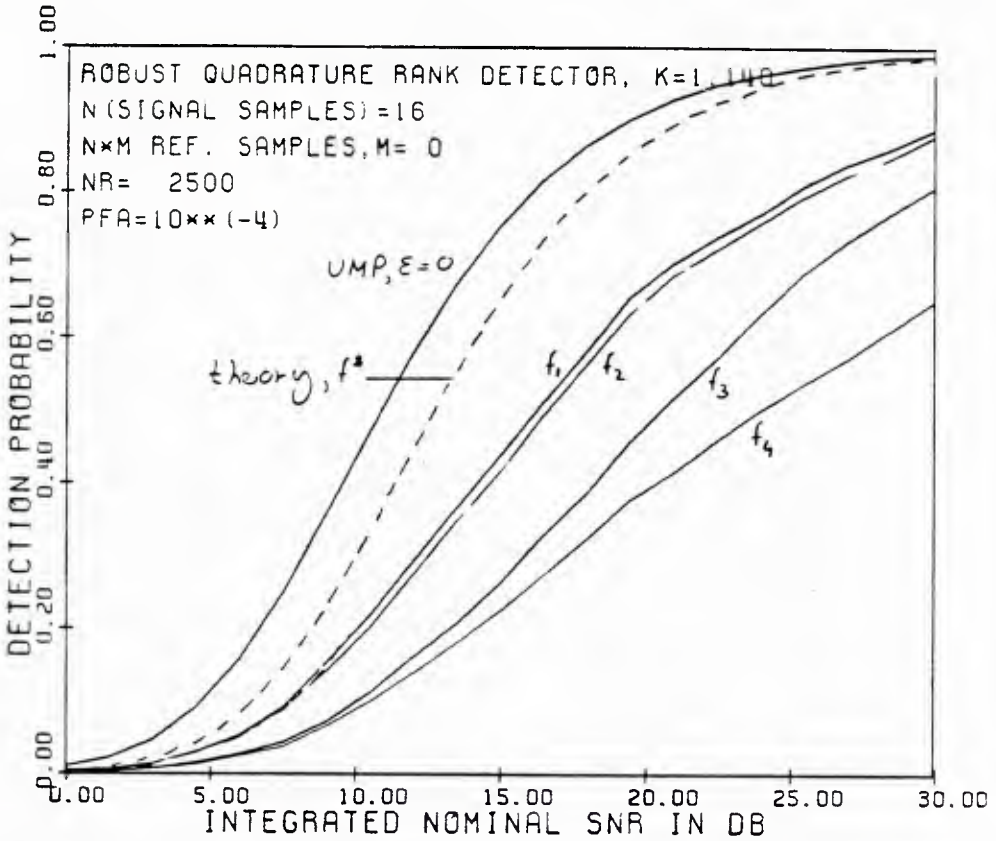


Fig. 6.7.7 Comparison of 1S-WX, VDW and RVDW tests, f_4 mixture noise, $n = 16$.

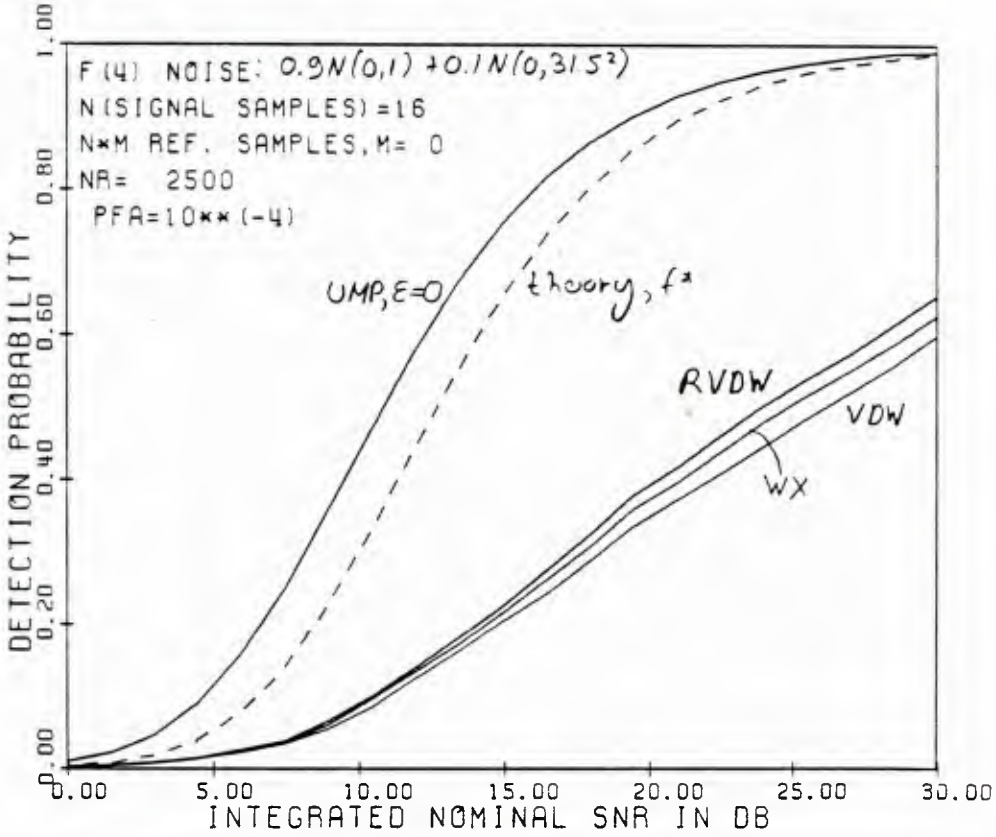


Fig. 6.7.8 1S-RVDW test, f_1 to f_4 noises, n =32.

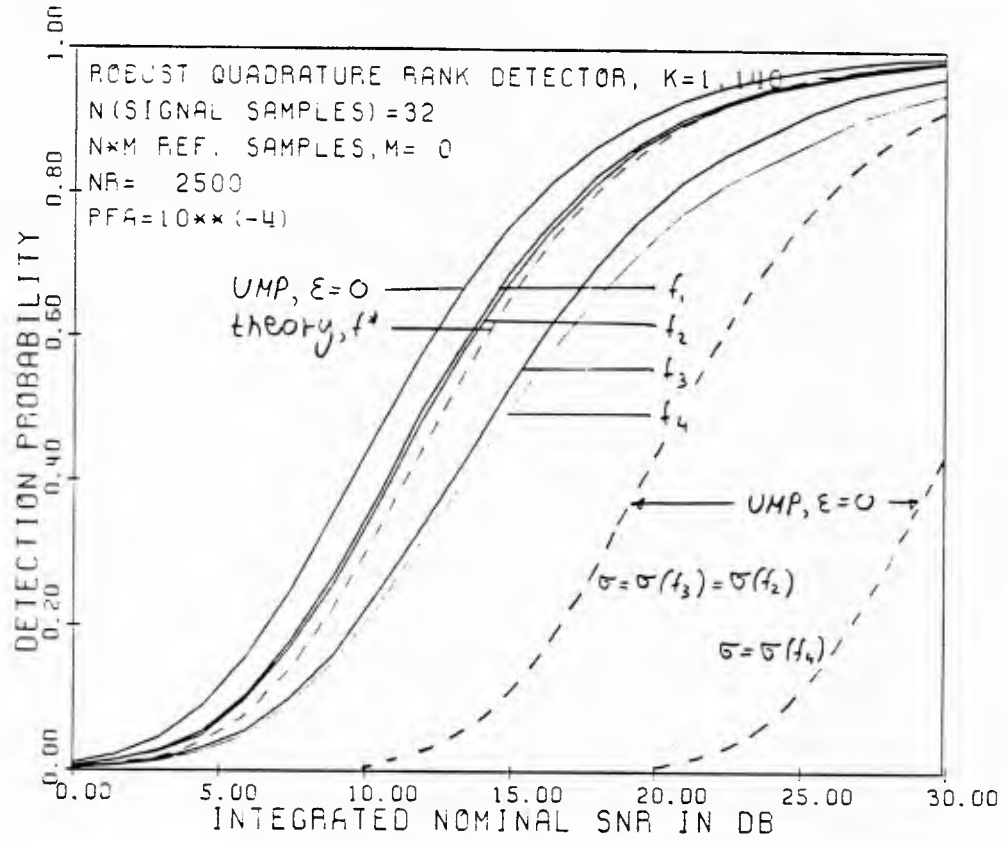


Fig. 6.7.9 1S-RVDW test, f_1 to f_4 noises, n =32.

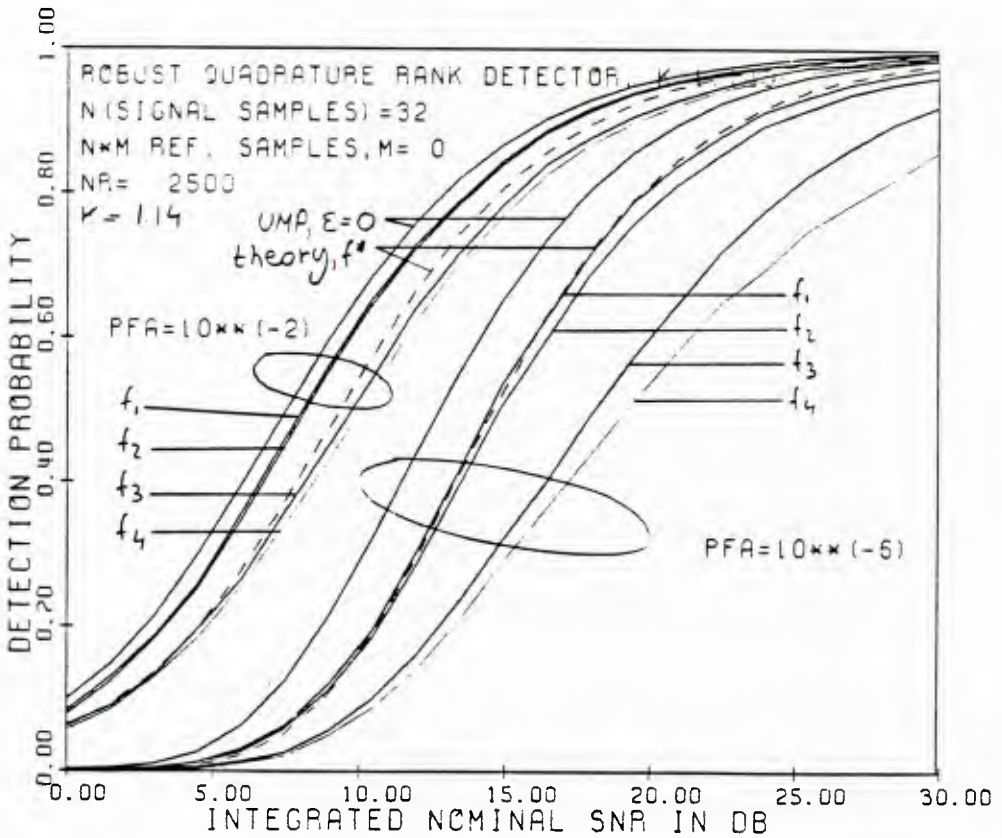


Fig. 6.7.10 2S-RVDW test, f_1 to f_4 noises, $n=32$, $M=2$

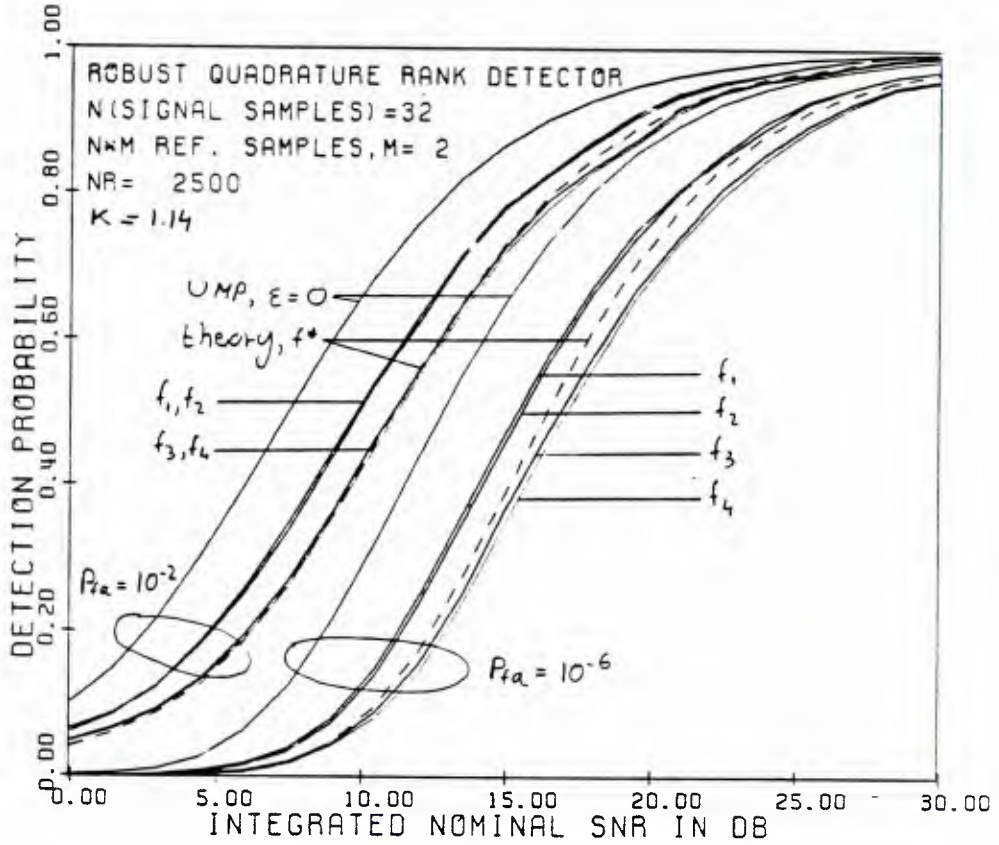


Fig. 6.7.11 Comparison of 1S-WX, VDW and RVDW tests, f_3 mixture noise, $n=32$.

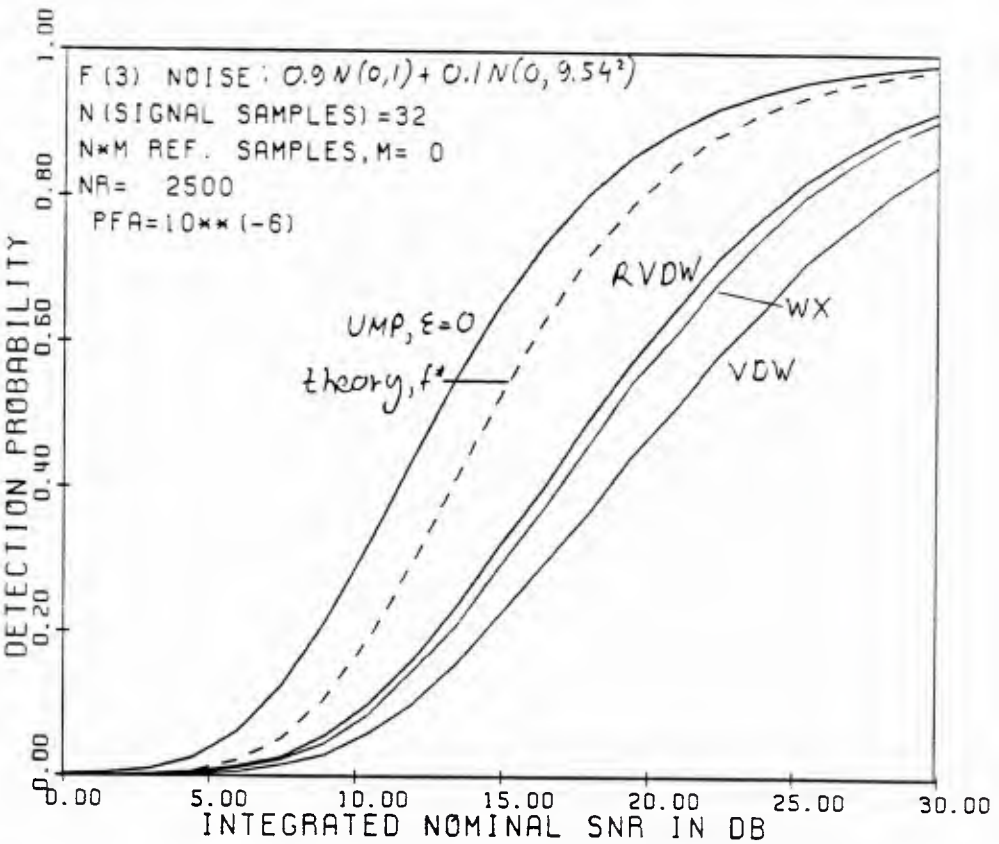


Fig. 6.7.12 Comparison of 1S-WX, VDW and RVDW tests, f_4 mixture noise, $n=32$.

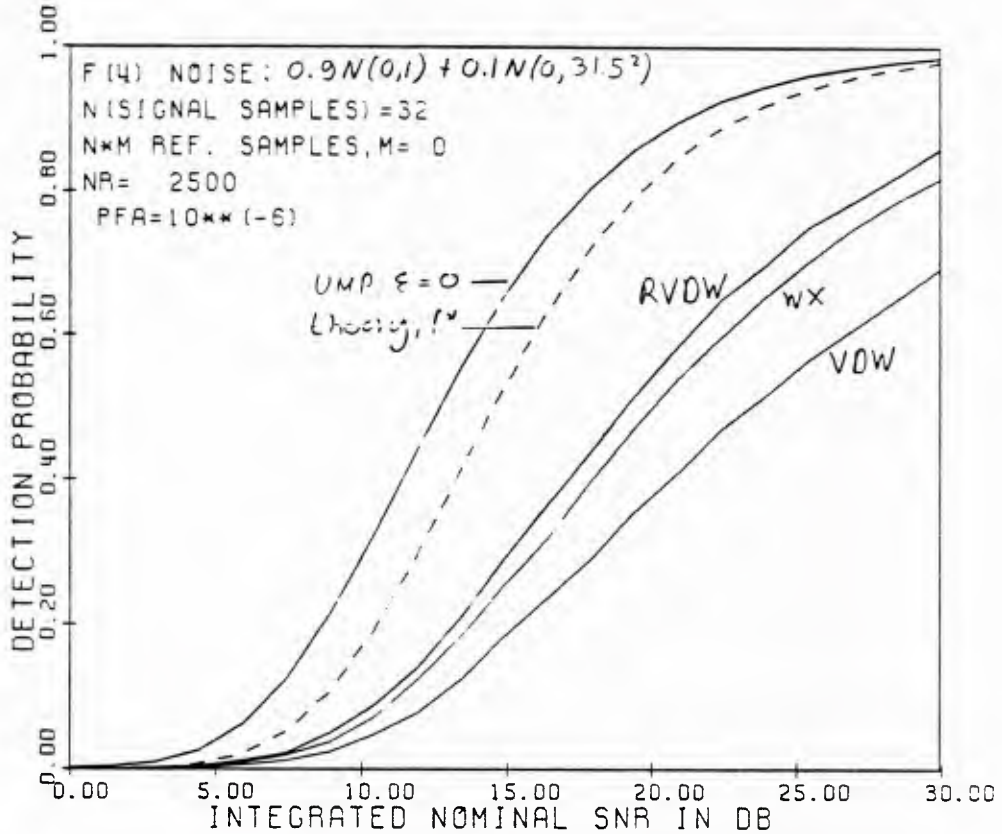


Fig. 6.7.13 1S-RVDW test, f_1 and f_4– f_6 noises, $n=32$, $k=0.862$.

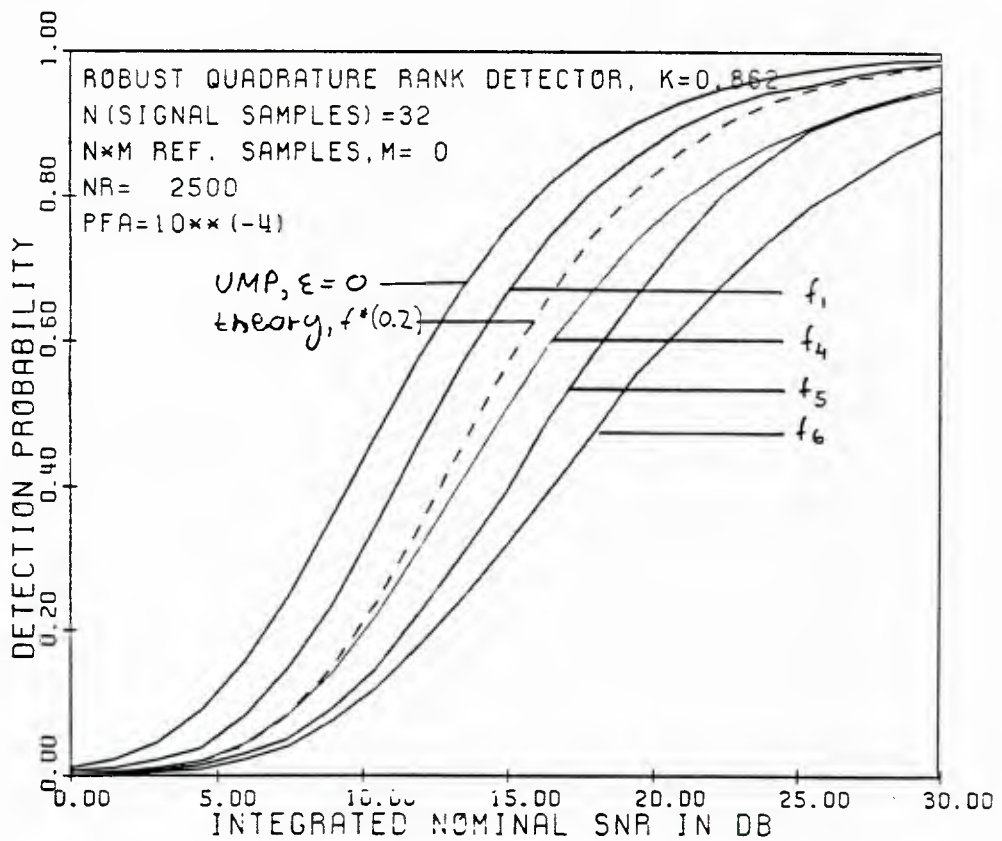


Fig. 6.7.14 Comparison of 1S-WX, VDW and RVDW tests, Gaussian noise, $n = 32$.

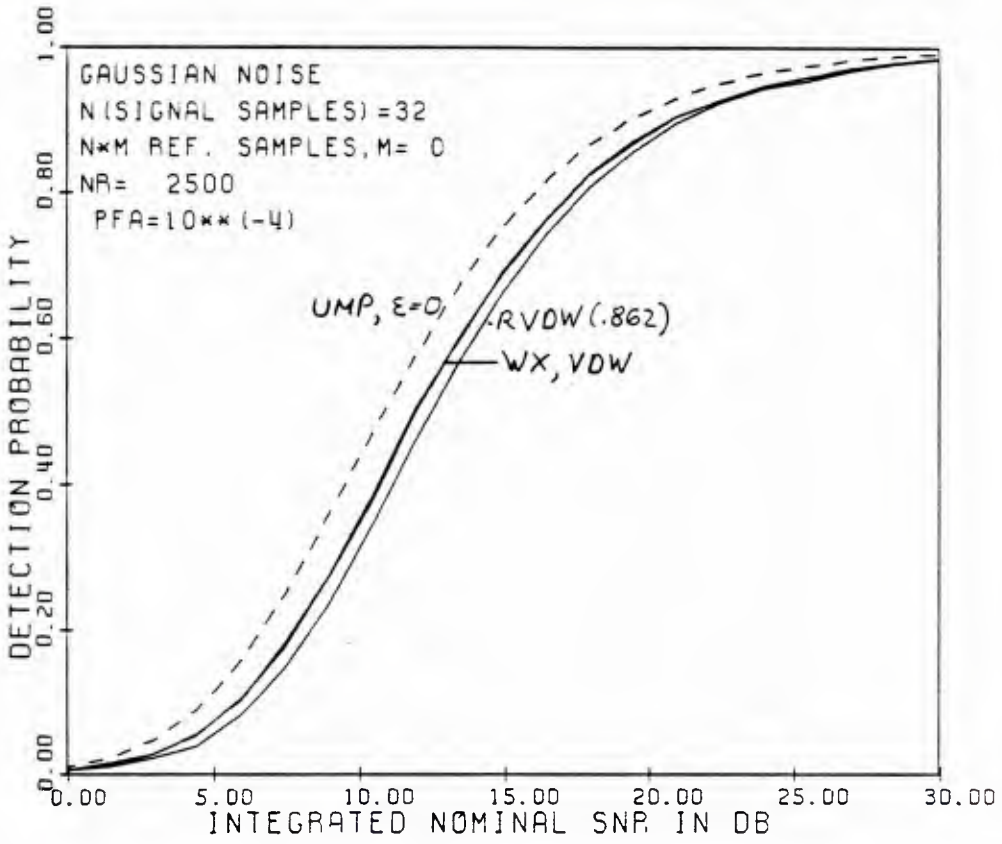


Fig. 6.7.15 Comparison of 1S-WX, VDW and RVDW tests, f_5 mixture noise, $n = 32$.

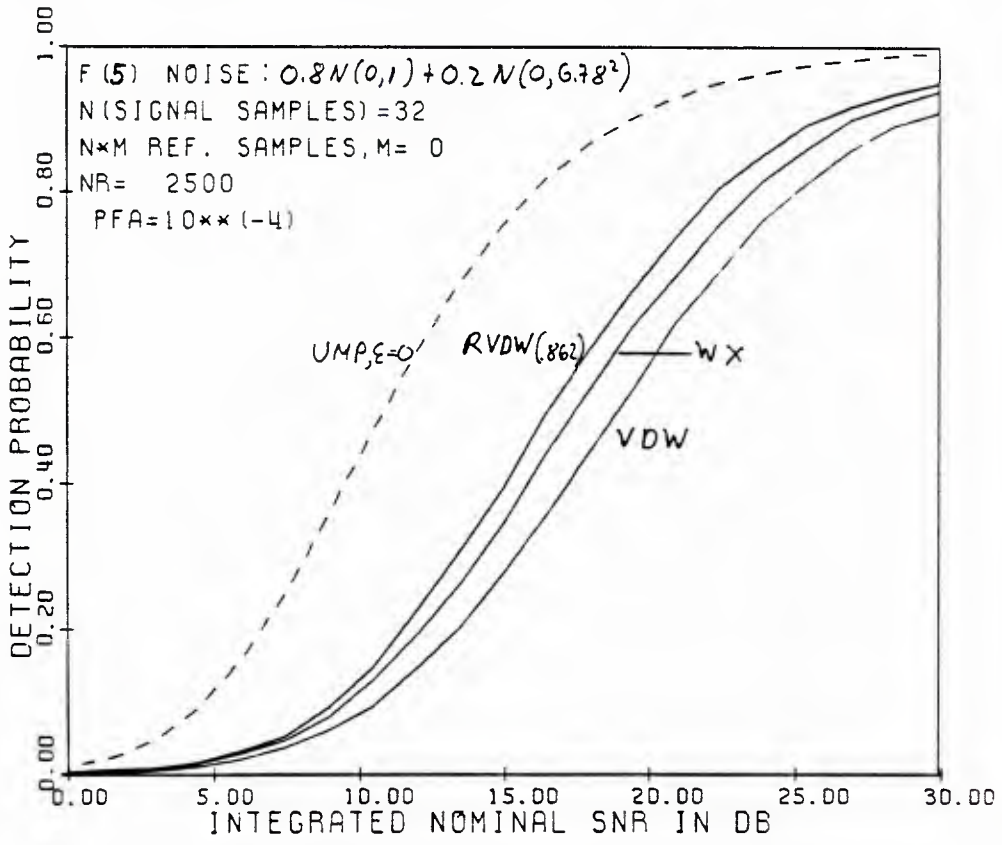


Fig. 6.7.16 Comparison of 1S-WX, VDW and RVDW tests, f_6 mixture noise, $n = 32$.

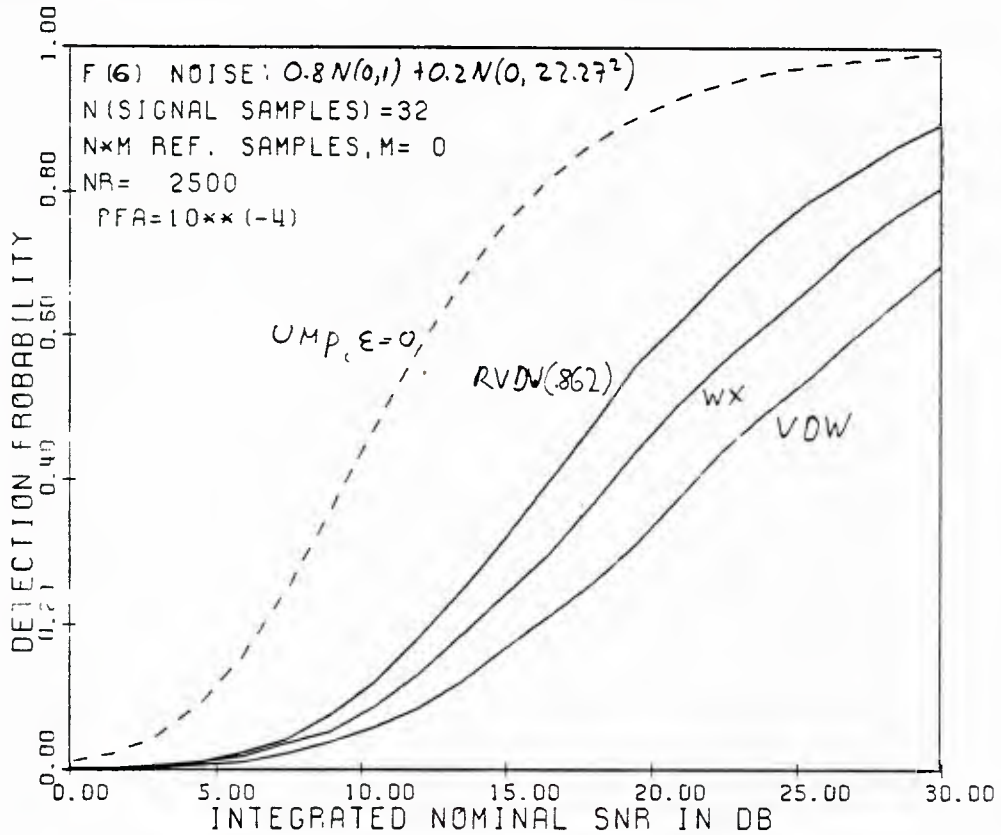


Fig. 6.7.17 1S-RVDW test, f_1 to f_4 noises, $n = 64$.

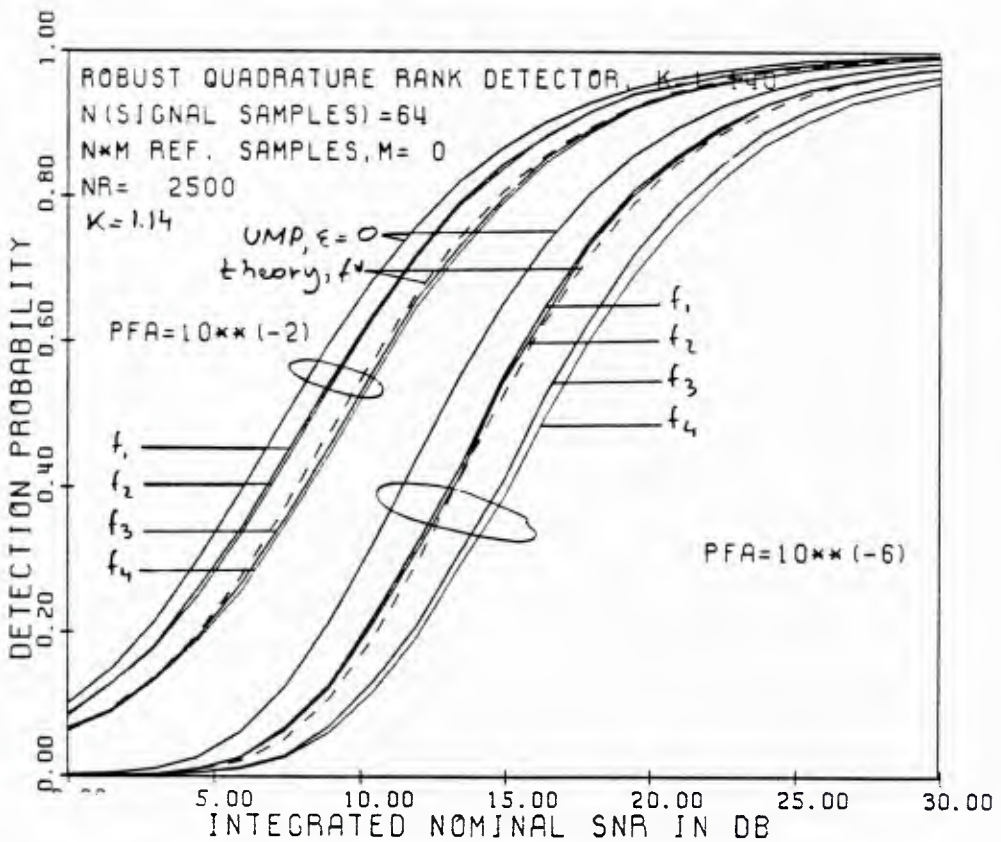


Fig. 6.7.18 1S-RVDW test, f_1, f_4, f_7 and f_8 noises, $n=64$.

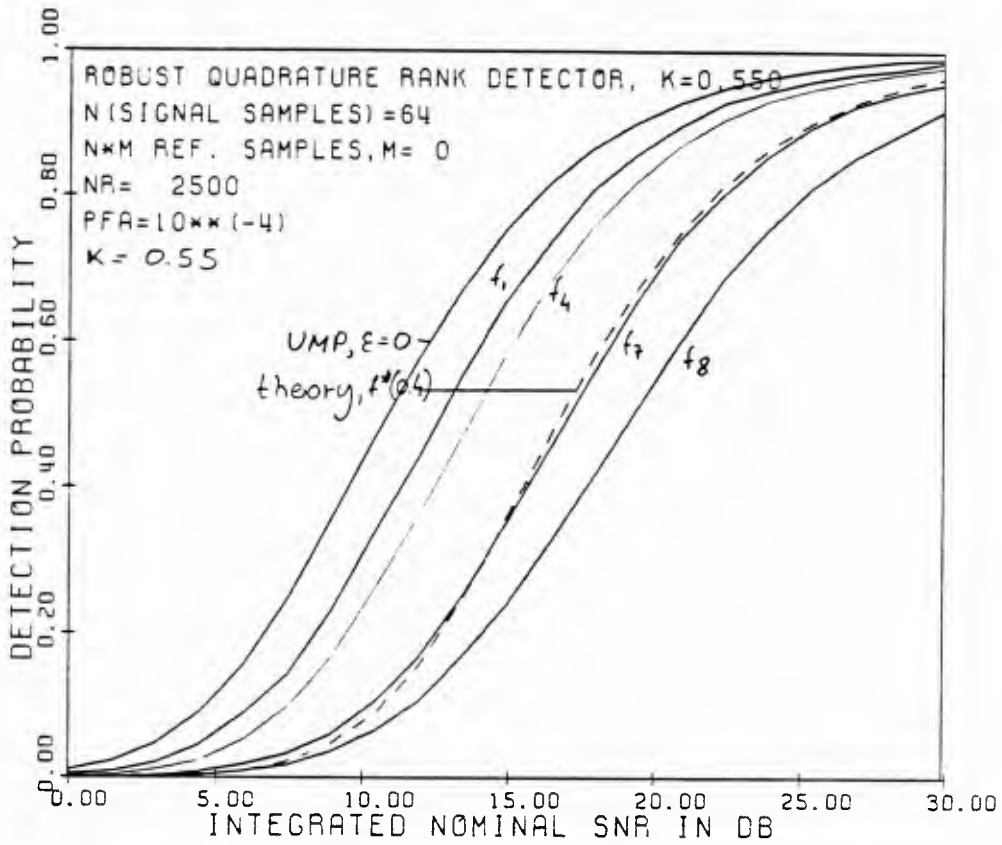


Fig. 6.7.19 Comparison of 1S-WX, VDW and RVDW tests, f_7 mixture noise, $n=64$.

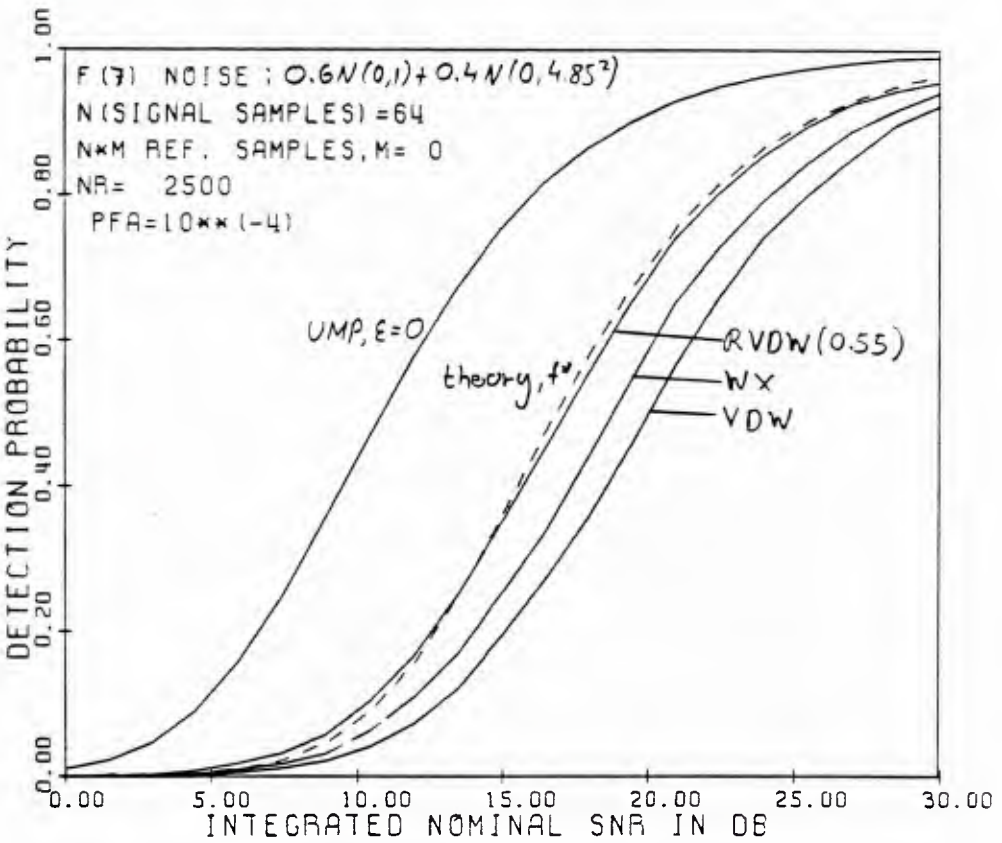


Fig. 6.7.20 Comparison of 1S-WX, VDW and RVDW tests, f_8 mixture noise, n =64.

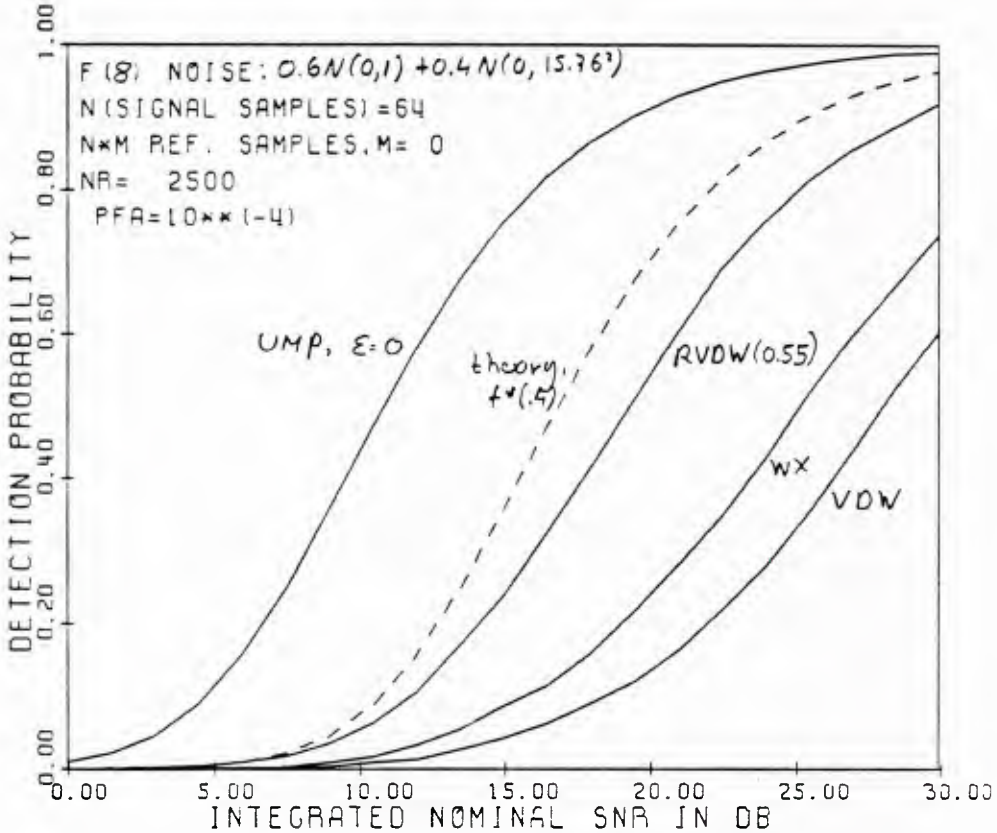


Fig. 6.7.21 1S-RVDW Doppler test, f_1 to f_4 noises, n =32, Δf =0.0.

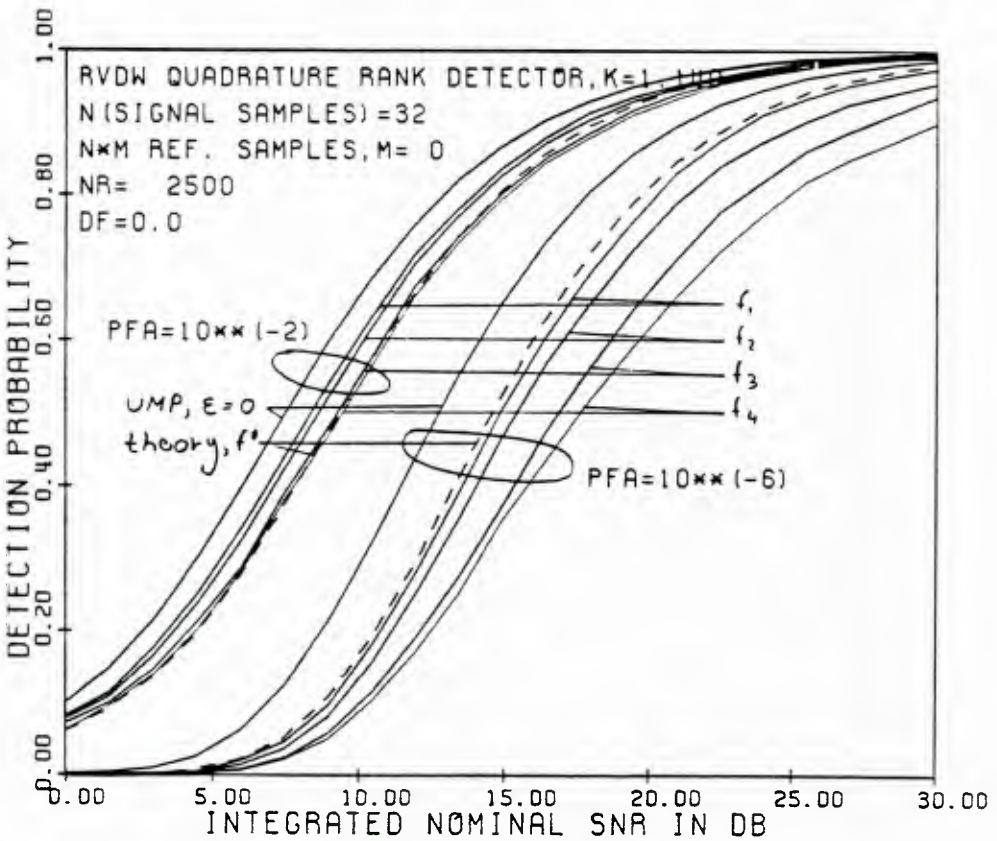


Fig. 6.7.22 1S-RVDW Doppler test, f_1 to f_4 noises, $n=32$, $\Delta f=0.25$.

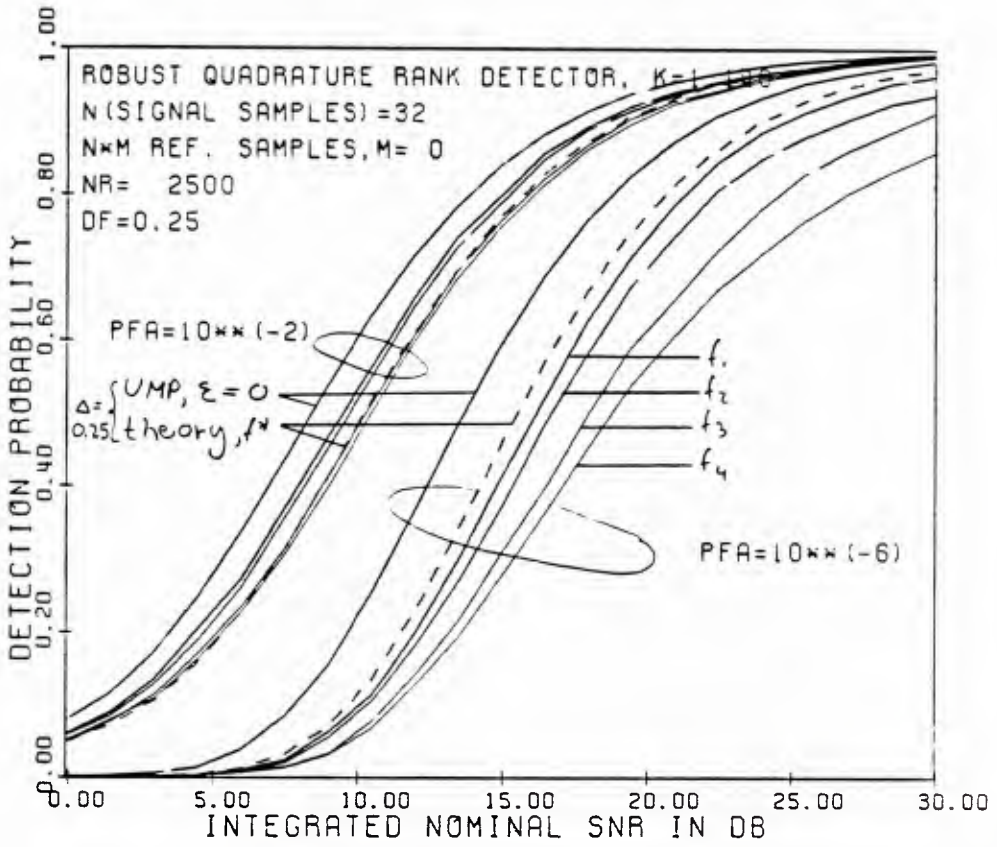


Fig. 6.7.23 1S-RVDW Doppler test, f_1 to f_4 noises, $n=32$, $\Delta f=0.50$.

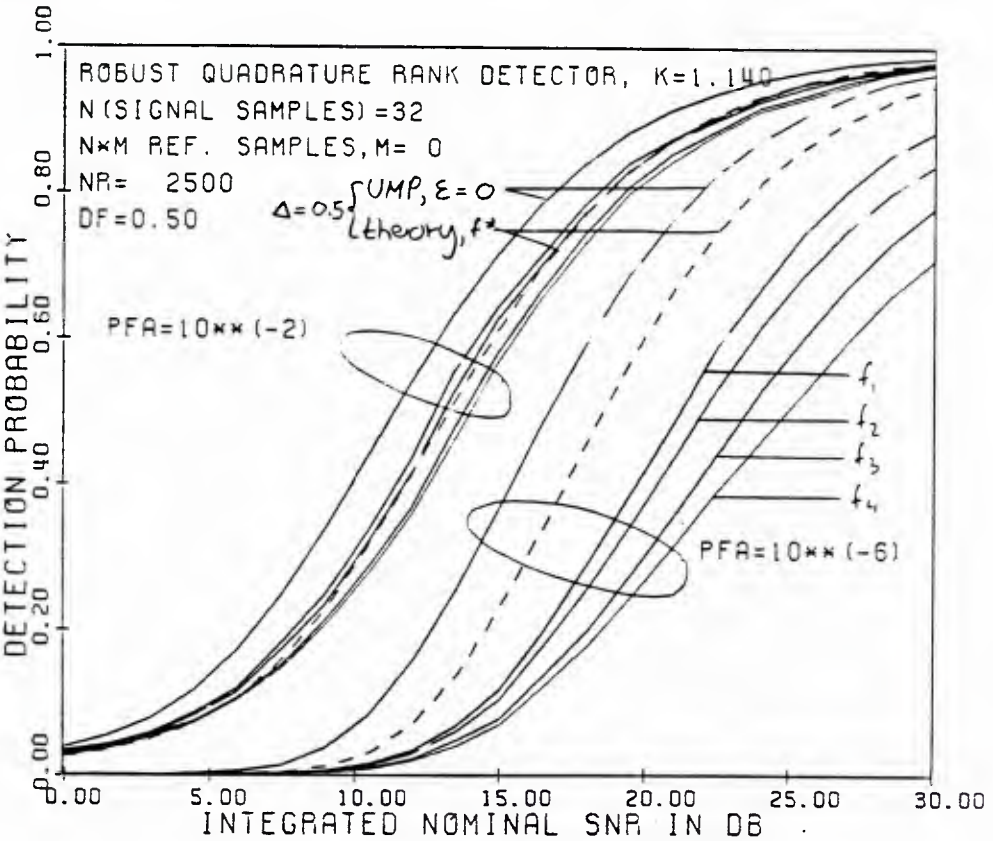


Fig. 6.7.24 2S-RVDW test, f_1 to f_4 noises, $n=16$, $M=4$

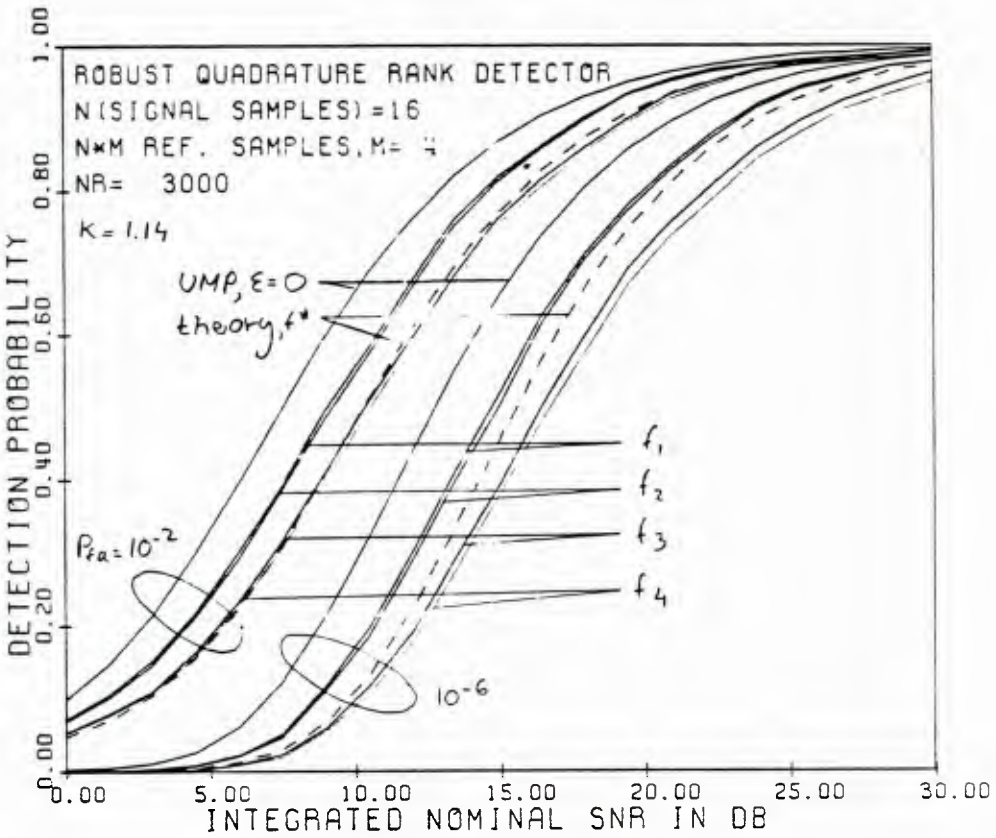
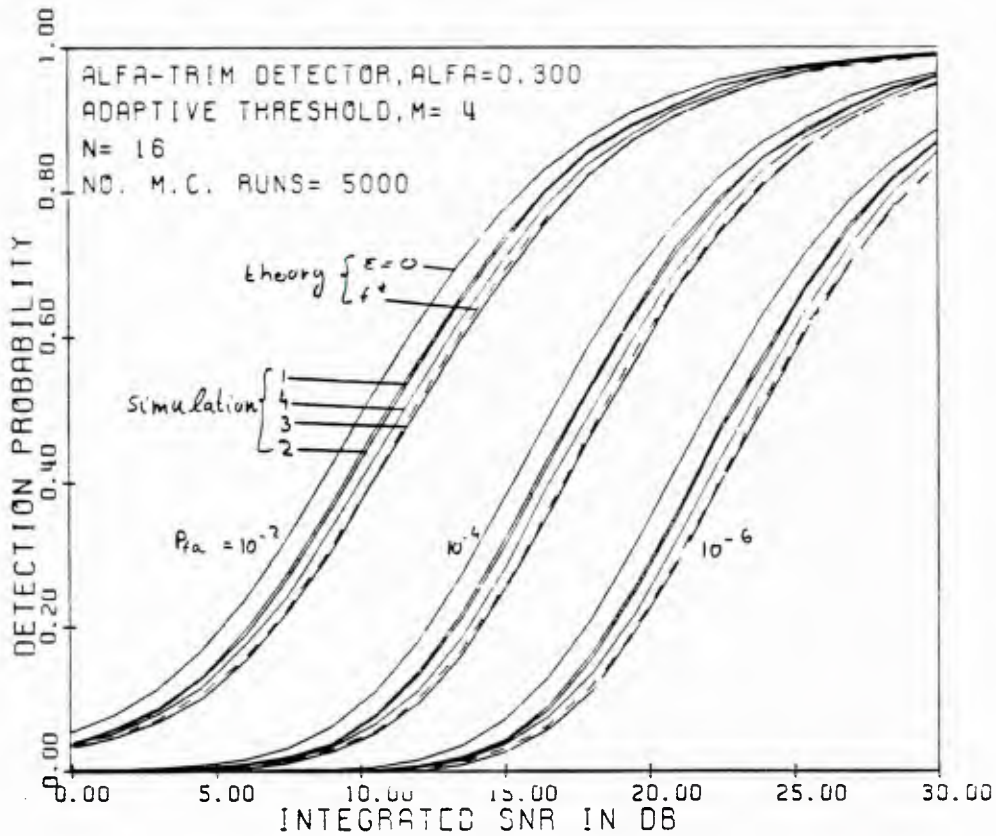


Fig. 6.7.25 SW-αtrimmed test, f_1 to f_4 noises, $n=16$, $M=4$



6.8 Overview and Comparisons of Different Robust Detectors

In Chapters 4 through 6, different algorithms for robust detection of fading narrowband signals have been developed. They vary in implementation, complexity and performance; thus, a comparison among them is appropriate.

Clearly, the SSQME test of Chap. 4 is the least attractive, and should be used only when the variance of the nominal noise is stationary and known precisely. Even then, the slightly more complicated sliding-window (SW) version of Section 5.2 will perform better for P_d values around 0.5.

With the other tests, none is uniformly best, and the choice would depend, mostly, on the environmental conditions. The major factors that have to be considered are the availability of noise reference samples (or the validity of the assumption that the target and reference samples are identically distributed under H_0), and how reliable is the assumption of independent samples. Since even under i.i.d. conditions the exact preformance depends on many factors (n, M, the desired P_{fa} and P_d, the maximum ϵ), only qualitative conclusions will be drawn here.

Complexity

Implementation of the SW or one-sample (1S) tests with M-estimators require, in general, more operations than with L- or R- statistics. The complexity is roughly the same for the SW and the 1S tests (either with the R- statistics or with the MAD scale estimator, Section 5.2), if the need for two orderings of the data in the 1S-SSQME is not counted. The two-sample (2S) R-test requires considerably more computations, and is more sensitive than the 1S R- (in an *homogeneous* environment) only when $n < 32$. Substantial reduction of complexity is possible with the single-sweep (SS) R- Doppler test (Section 6.6.2), which is the only one that preserves the $O(n \log n)$ of the conventional periodogram (FFT) test. (Its small sample performance, though, has not been simulated yet).

Compatibility with non-homogeneous environment.

The analysis in this thesis has been restricted to the cases where the target and reference samples are identically distributed. But, heuristic assessment can be made for spatial non-stationarity. In radar systems, the major sources for it are the following: a) The noise power varies as a function of the type of terrain from which the clutter echoes are reflected. This factor is less significant if the terrain is fairly uniform (e.g., non urban areas) and the range extent of the reference window is relatively short (say, 0.5 km). High range resolution (say, 50 meters), and reasonably small M (say, $M=10$) are sufficient. Moreover, in any reasonable coherent radar processing, clutter cancellation (i.e., whitening by means of notch filtering the clutter spectral peak) is performed prior to other operations. Properly designed, the clutter residues are at about the thermal and quantization noise levels, and consequently, the influence of non-stationarity is reduced. b) "Edges". Sharp increases in the clutter levels are encountered on the boundary of different terrains (e.g., when mountain ranges touch the seashore line). The major problem with edges, is that they cause a false alarm whenever the window sweeps over the edge. Again, clutter prewhitening reduces this phenomena. c) An echo from another target within the reference window, is common in dense target situations. Closely spaced targets (range separation smaller than M) cause a capture phenomena; even when the SNR is very large for both targets, and they are roughly equal, P_d is often limited to very small values.

The various 1S tests are clearly immune to any non-homogeneity, as no reference is used. The sensitivity of the 1S R- test is lower by several dB's than that of the test with α–trimmed and MAD estimators when $n<32$ and $\epsilon\leq 0.1$. However, this disadvantage is compensated by the DF-CFAR operation of the first for all sample sizes, and even for nonsymmetrical noise density. Moreover, for very large contaminations (say, $\epsilon>0.3$) and small n, the R- based 1S test has better sensitivity. (Recall, however, that phase modulation of the transmitted pulse train is required for the R- test).

When is the SW test with α–trimmed estimator better? Its sensitivity in a homogeneous environment is considerably better (for $n=16$, $P_{fa}<10^{-4}$, $P_d\cong 0.9$) provided $M\gtrsim 6$. With this window size, however, it becomes more susceptible in a non-homogeneous environment to the above mentioned phenomena. When the type a) non-homogeneity is not a serious problem, it is relatively simple to protect the SW test against the undesired effects of the b) and c) types. Since the behavior of the robustified SW test in a contaminated noise has been shown to be very similar to that of the linear SW test in Gaussian noise (where the only difference is the modification of the effective SNR), even for small sample sizes, the same techniques that have been applied before are appropriate [19-20]. The cure against edges is to modify the adaptive threshold $W_\alpha(\mathbf{u}, \mathbf{v})$ in Eq. (5.5) such that the sample averages of the leading and the trailing halfs of the window are computed, and then the greatest of the two is taken as the adaptive threshold, see [19]. The capture problem caused by closely separated targets is treated by censoring from the reference window the cell of largest magnitude (or the m largest cells if m targets are expected to be within the window). Censoring and "Greatest-of" detection are easily combined to protect against both types b) and c) of non-homogeneity, see [20]. (The threshold multiplier $t(M)$ is different in the combined algorithm, but it is possible to compute it with the techniques of [20]).

Detection in dependent noise

Throughout the thesis, the noise samples have been assumed to be statistically independent. Recently, several papers have treated robust detection for serially dependent noise samples. Poor [75], following Portnoy's[74] work on robust estimation with dependent noise, considered weak dependency resulting from a first order moving average model, with parameter $\rho \to 0$. His robust M-detector utilizes a linearly corrected version of Huber's soft limiter, and is capable of restoring the $O(\rho)$ sensitivity loss which the usual M-detector suffers in the assumed moving average noise. Moustakides and Thomas [76] looked at a more general structure for dependency, namely at ϕ–mixing noise

sequences. They have proved a minimax result on the efficacy, and an upper bound on P_{fa}, for the class of limiter-correlator detectors, and for ϕ–mixing noise with an ϵ–mixture marginal distribution. Their test utilizes a null- zone modification of Huber's limiter. Although these results are theoretically important and similar structures could be sought for the signal models in this thesis, it can be argued that this would not be the best direction to proceed.

First, when dependency (in radar detection) is a problem, it usually is a *strong* dependency [1]. It is well known that the optimal detector for dependent Gaussian noise first prewhitens the noise, and then correlates it with the known signal. Therefore, the above robust tests which are all based on *memoryless* operations, are far from being the globally optimal structures. (It is easy to demonstrate that a memoryless detection when the noise is Gaussian but strongly correlated, can cause huge degradation). Second, the modifications on the non-linearity require exact knowledge of some parameters in the dependency structure; in particular, the mixing parameters $\{\gamma_n\}$ in [76, Eq. (4)] are very difficult to determine. This precise knowledge seems to contradict the spirit and purpose of robust statistical processing.

On an intuitive basis, it seems that the natural way to attack the problem is first to try to prewhiten the input noise (utilizing adaptive techniques of modern spectral estimation [78] if the autocorrelation is unknown), and then to apply one of the robust tests

[1] In radar detection, when the noise is dominated by ground or precipitation clutter, there is a very strong dependency among the input samples. The power spectrum (of the discrete autocorrelation) is usually composed of a relatively flat pedestal, representing the thermal noise and clutter echoes the are received through the antenna sidelobes, and a very high and narrow peak, that corresponds to the clutter reflections through the antenna mainbeam. (When weather and other aerial clutter are present, they produce another high and somewhat wider spectral peak, that is shifted by the average Doppler frequency of the producing clouds from the ground velocity). Prewhitening of this colored spectra type is pretty simple in the ground clutter case; the center frequency of the clutter peak is tracked by phase lock loops, and its width is essentially known as it is related to the antennas beamwidth and its scan frequency. Thus, prewhitening is done by notch-filtering ("MTI" in radar terminology). When aerial clutter is also present, it can be whitened by cascading the ground clutter canceller with another phase lock loop/notch filter combination. Alternatively, modern adaptive techniques of spectral estimation can be used to estimate the spectrum and to control the parameters of the prewhitening filter [78].

for independent noise. Since prewhitening is a reversible operation, an optimal procedure following it would not involve any loss, see [14, Part I, p. 289]. Although white but non-Gaussian noise is not necessarily independent, this approach might be acceptable. (For estimation of a location parameter, Martin [77] has recently suggested that this is the "proper" approach to achieve efficiency under dependence; he has also pointed that this is the proper way to design robust tests).

Now, there remains the question which of the structures that have been studied in this thesis is more appropriate. The prime objective is to achieve the DF-CFAR performance, for any dependence scheme. Under certain conditions, the robust M- and L-estimates as well as the R-statistics are expected to be asymptotically Gaussian distributed, but with a variance different from the i.i.d. case. The variance will depend (in addition to other factors) on the dependence structure. Proof for M-estimates with either a first order autoregression or a first order moving average scheme was given by Portnoy [74], see also [77]. Wolff et al. proved it for linear test statistics with any finite order autoregression, see [79] and the other references given there.

From all the normalization structures studied in this thesis, the SW class of tests will assure DF-CFAR as long as the in-phase and quadrature statistics are asymptotically Gaussian. Notice that the various R-tests are not CFAR anymore under dependence, but a SW version of the 1S R-test is easy to construct. Simply, the fixed threshold in the 1S version of Eq.(6.3.4) is replaced by the sample mean, taken over the reference window, of the same quadrature statistic which is computed for the cell under test. "Studentizing" the 2S or SS rank tests is also possible, following the ideas in [79], but demands a substantial increase in the complexity. (Basically, an estimator of the variance of the "main" test statistic is needed to set the threshold). Since even the i.i.d. 2S-R test is more complex than the 1S, and is more sensitive only for quite small sample sizes, it seems that the 1S is a better choice, especially under dependence. Sacrificing the computation efficiency of the SS R- Doppler test, however, is a disadvantage.

Although a deeper and thorough analysis of the suggested scheme is beyond the scope of this work, a simulation of a single example was carried to demonstrate the validity. We assumed that following the prewhitening the noise can be modeled as a first order moving average process as in [75], where the i.i.d. samples from which the moving average noise is derived have an $\epsilon-$ mixture distribution. The moving average parameter was ρ=0.1, and the test was the SW (n=16, M=4) with α–trimmed estimators. It was found that the tails of the distribution of the test statistic under H_0 were somewhat longer; however, increasing the trimming ratio to α=0.4 kept the P_{fa} very close to the asymptotic prediction. (See Fig. 6.7.26.) Actually, the curves are very similar to Fig. 5.9. The P_d curves also turned out to be very similar to those of Figs. 5.17-5.19, with the same insensitivity to the marginal distribution; see Fig. 6.7.27. Here, the dashed curves correspond to the computed P_d of the UMP test when the noise is a MA Gaussian with ρ=0.1; the other are for MA noises with the f_1-f_4 generating distributions of Section 5.4.

Naturally, the dependence causes a sensitivity loss compared to the i.i.d. case. This, however, is roughly identical with that suffered by the optimal detector for Gaussian noise with the same correlation matrix. The latter can be shown to multiply the SNR, $A^2/2$, by $\sum_{i,j} R_{ij}^{-1}$ instead of by n, where R is the correlation matrix. For the assumed model, this loss is, with ρ=0.1 and n=16, 13.6/16=0.7 dB, and the dashed curves in Fig. 6.7.27 are shifted by this amount from the ϵ=0 curves of Fig. 5.19. For the longer tailed distributions f_2-f_3, there is an additional sensitivity loss of about 0.8 - 1 dB. It is mostly attributed to the increased estimator's variance when α is raised from 0.3 to 0.4.

Fig. 6.7.26 P_{fa} of a SW- α–trimmed test, under first order MA dependence.

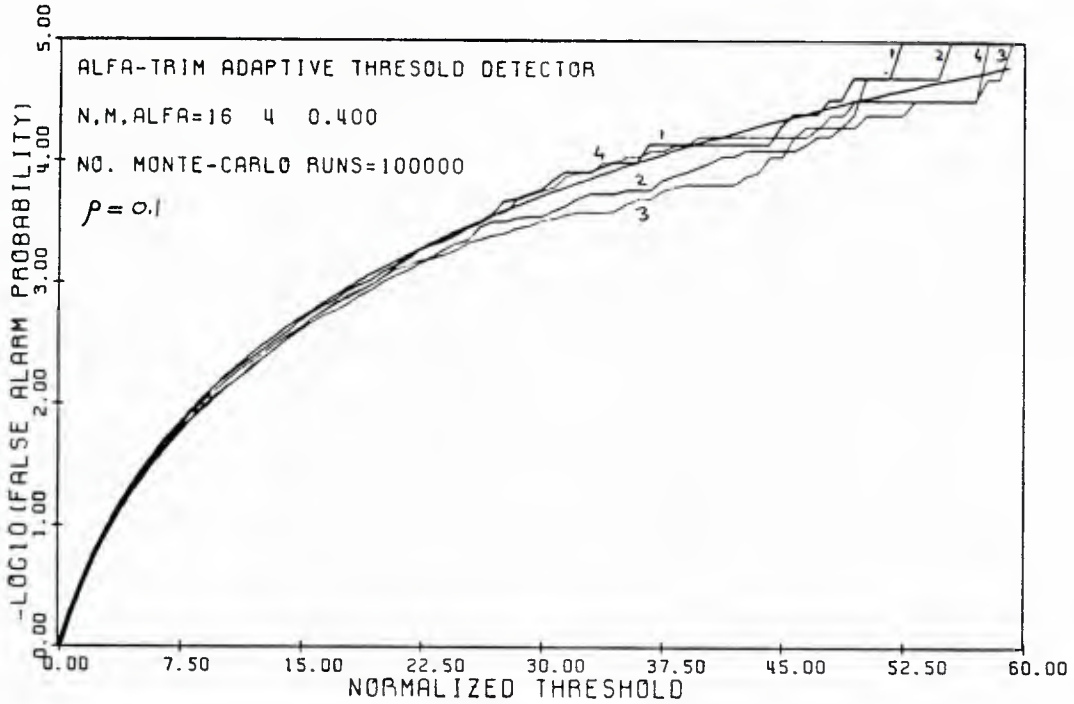
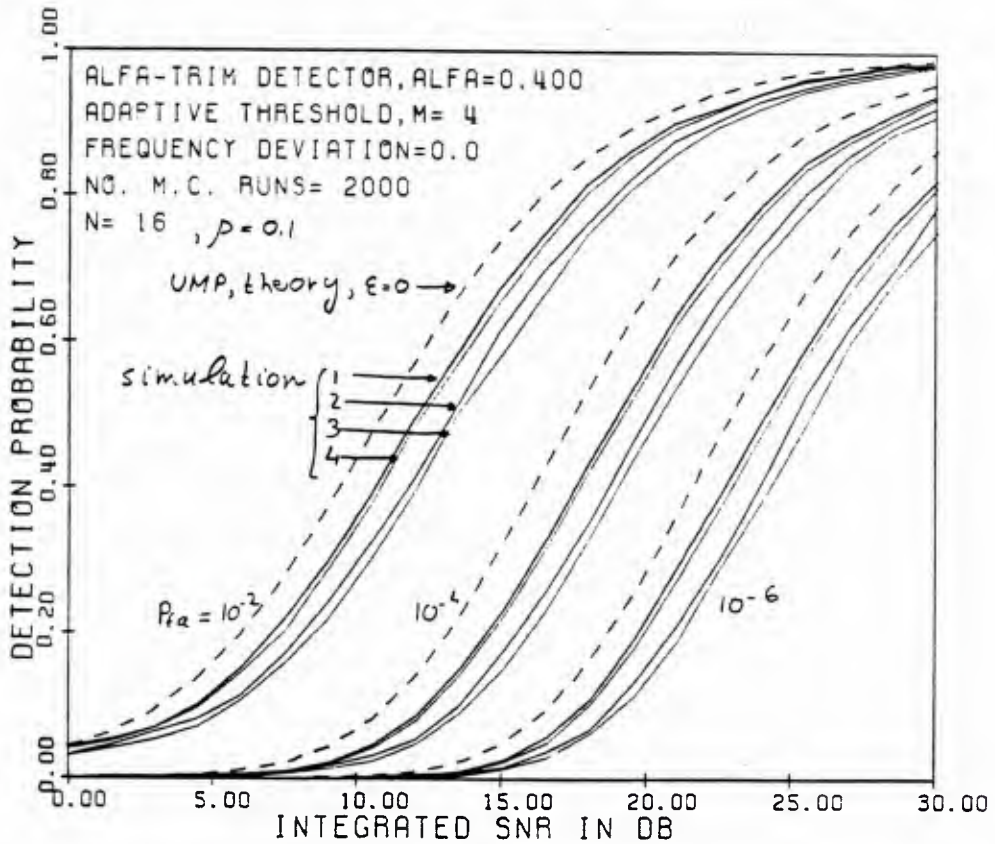


Fig. 6.7.27 P_d of a SW- α–trimmed test, under first order MA dependence.



7. ROBUST DIGITAL M-ARY COMMUNICATION RECEIVERS

7.1 Introduction

It appears that optimal digital communication theory is still mostly a theory of additive Gaussian noise. A survey of both classical and modern texts, cf. [65-68], does not reveal any significant results for non-Gaussian noises. This is true for both aspects of the communication problem: optimal demodulation and optimal signal selection. On the one hand, optimal digital demodulation is analytically tractable for arbitrary complicated signal sets received in additive colored Gaussian process. Moreover, explicit expressions or tight bounds on the error probability exist, and it is therefore possible to find optimal signal sets. On the other hand, information theory supplies us with a closed form expression for the channel capacity of the additive white Gaussian noise channel (AWGN), against which the efficiency of any signaling scheme can be compared.

The tractability of both aspects of the problem is attributed to the remarkable property of a Gaussian random process $n(t)$: any integral of the form $\int_a^b f(t)n(t)dt$ is a Gaussian random variable. Hence, infinite dimensional decision problems are made countable by such techniques as the Karhunen-Loe've expansion, projection on the basis set spanned by the signal set to obtain sufficient statistics, or whitening augmented by a reversibility argument. Non-Gaussian processes, on the other hand, pose formidable difficulties. For the optimal detection part of the problem, the receiver should maximize the posterior probability that a signal $s_m(t)$ has been transmitted. However, for cases of prime interest we do not have close form expressions for the posteriors. Consequently, the optimal signal set for given constraints can not be found. While numerical algorithms for computation of the channel capacity of continuous channels and source alphabets have been derived [67], we do not know of any closed form expressions for non-Gaussian cases. Hence, there is no baseline against which sub-optimal receivers could be compared.

As in binary hypothesis testing, we would expect that various communication environments consist of non-Gaussian noises. These include, to name a few: RFI or jamming in radio communication, other non-synchronized messages in multiple-access channels, under-water communication, etc. We thus might wonder why, in contrast to binary hypothesis problems, robust digital communication has not been studied so far. One explanation is the previously mentioned analytical difficulties. The second is that, perhaps, optimal communication is not the focus in practice. For instance, SNR's of 70 dB are typical in the American telephone network. Yet another explanation arises from information theory - for fixed signal energy and additive noise power, the capacity is the lowest for the AWGN channel - see Theorem 7.4.3. in [66]. Hence, if someone compares the performance of a given receiver only against the input SNR he would not be alarmed. As a consequence of Shannon's coding theorem, however, we know he could have done better had he utilized some knowledge on the non-Gaussian distribution (this will be demonstrated in the sequel). Finally, some ad-hoc devices in the receiver, which usually are not designed as an outcome of statistical analysis (e.g., squelch circuitry in FM receivers, Dolby in audio tape decks, etc.) might combat successfully some of the non-Gaussian interferences.

In this chapter we attempt a modest first step towards robust digital communication for additive, white non-Gaussian (AWNGN) channels. To circumvent the analytical difficulties we study only the discrete time problem, and we limit the scope to the simplest modification of the discrete time white Gaussian process, i.e., we modify only the first order distribution. Within this framework, we show that there is a close relationship to the problem of robust estimation of a location parameter by means of M-estimators. Further, by keeping the structure of the bank of correlators receiver, but replacing the correlators by limiter-correlators, a minimax error-probability is obtained when the limiter is Huber's soft limiter function. This is true for an arbitrary signal set. Similar results are obtained for noncoherent communication, where the receiver is

robustified by inserting limiter-correlators in the in-phase and quadrature channels of the quadrature square-law receiver.

When it is desired to reduce the sample size, we propose a similar structure with generalized M-estimators replacing the limiter-correlators. Specific important cases (PAM, PSK and QAM) yield identical asymptotic results. The same is shown true for the orthogonal and uniform-distance signal sets, provided we replace the one-dimensional M-estimators by the solutions of an equivalent robust regression problem. Based on our Monte-Carlo experience with the binary problems in previous chapters, we suspect that the latter implementation is superior to the first.

7.2 Robust M-Ary Coherent Receivers: Limiter-Correlator Implementation

In the M-ary digital communication problem, at each time interval $t \in [0, T]$, the transmitter sends one of the signals $As_m(t)$, corresponding to one of equally probable M message symbols $m \in \{0,1,...,M\}$. The signals are characterized by their energies ϵ_i [1] and the normalized inner products ρ_{ij}

$$\int_0^T s_i(t)s_j(t)dt = T\sqrt{\epsilon_i \epsilon_j}\,\rho_{ij} \tag{7.1}$$

where $\rho_{ii} = 1$ and $|\rho_{ij}| \leq 1$. Except for pulse amplitude modulation (PAM), we assume $\epsilon_i = \epsilon \forall i$.

The coherently received signal at the output of an AWNGN waveform channel is

$$r(t) = As_m(t) + w(t) \tag{7.2}$$

where $w(t)$ is a band-limited, zero-mean, wide-sense stationary (w.s.s.) random process, not necessarily Gaussian, with spectral density $S_w(f) = CN_0/2$ for $|f| \leq B$ and 0 otherwise. We consider the "nominal" power spectral density $N_0/2$ as known, but $C > 0$ as unknown. In particular, we will be interested in heavily tailed marginal noise distribu-

[1] In our notation, $s_i(t)$ and ϵ_i are dimensionless. The physical energy is thus given by $A^2 T \epsilon_i$.

tions, for which $C >> 1$ ($C = 1$ for Gaussian). We will take $n = 2BT >> 1$; this ensures that the received signal is essentially undistorted as well as that the noise process approaches a white noise in the limit.

In order to avoid some very tough analytical difficulties that are inherent to any detection problem with continuous time, non-Gaussian processes, we will limit our study only to discrete time processing [2].

The received waveform is time sampled at the Nyquist rate $2B = 1/\Delta t$, yielding $r_i = r(i\Delta t)$, $w_i = w(i\Delta t)$ and $s_{mi} = s_m(i\Delta t)$, $i = 1,....,n = 2BT$. Obviously, the noise samples are uncorrelated. We make the following crucial assumption on the class of noise processes $\{w(t)\}$ to be considered hereafter:

If the $\{w_i\}$ are uncorrrelated , they are statistically independent (p)

It is justified to ask if the postulated class contain any other process beside the Gaussian. This question has been positively answered by Martin and Schwartz [69]. They have constructed the so called quasi-mixture Gaussian process (QMGP), which possess the above property if the quasi-mixture is generated from two independent white Gaussian processes; the marginal distributions are simply ϵ–mixtures of two Gaussian p.d.f.'s. [3] While the definition and construction of the QMGP might seem somewhat artificial and non physical, any other definition of a w.s.s. non-Gaussian process that does not obey (p) is bound to involve formidable analytical difficulties. Hence, even though a more realistic characterization might be desired, the one we have postulated

[2] Admittedly, there are two major drawbacks. a) In some cases, it is simpler and more economical to implement the digital communication receiver with analog processing, e.g., the matched filter in a PSK receiver is commonly implemented by multiplying (with diodes) the input signal by a sinusoid at the carrier frequency, and then integrating the product by a capacitor-operational amplifier circuit. b) Even when the number of samples is very large, the performance might be inferior to that achieved by the optimal analog processing. Although this is not the case with Gaussian noise and the bank of matched filters receiver, we can not claim it in general without a deeper analysis.

[3] While it does not seem possible to realize a continuous QMGP exactly, arbitrary close approximations can be realized by switching between two independent Gaussian processes, where the switching rule is based on a nearly white Poisson process. At each time increment of the Poisson process, the first Gaussian process is selected with probability ϵ, see [69].

will serve at least as a modest first step towards robustification of digital receivers against deviations from the predominant AWGN channel assumption.

Without loss of generality, the receiver first normalizes the input by the square root of the "nominal" power: $x_{ni} = r_i / (N_0 B)^{1/2}$. We now have a multiple-hypothesis detection problem, where on the m^{th} hypothesis

$$H_m : \quad x_{ni} = a_n s_{mi} + n_i \quad , i = 1,....,n \quad m = 1,....,M \tag{7.3}$$

Here, $a_n = A / (N_0 B)^{1/2} = O(n^{-1/2})$, and the n_i are i.i.d., zero-mean, with variance C and a symmetric p.d.f. f . As B and $n \to \infty$, (7.1) is replaced by

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} s_{mi} s_{li} = \sqrt{\epsilon_m \epsilon_l} \rho_{lm} \tag{7.4}$$

We restrict our attention to a class of decision rules which capture the structure of the (discrete time) optimal receiver for the AWGN channel, but we replace the linear correlation by a nonlinear one. Namely, we decide in favor of H_m if $d_m = \max_{i \in \{1,..,M\}} d_i$,

$$d_m = \frac{1}{n} \sum_{i=1}^{n} l(x_{ni}) s_{mi} \quad m = 1,...,M \tag{7.5}$$

and $l(\cdot)$ satisfies the same regularity conditions as in Chapter 4. [4] Heuristically, it seems plausible that if $l(\cdot)$ is some sort of a limiter, the effect of heavy tailed noise will be reduced, but some performance loss will be incurred if the noise is Gaussian. We will show that if the marginal noise p.d.f. f belongs to a class of distributions for which a minimax M-estimator of location exists, then a minimax result on the error probability is obtained *within* this class of decision rules.

[4] Notice that for any noise process, the optimal Bayesian receiver computes posterior probabilities and decides in favor of the largest, but we can not claim that the decision variables of Eq. (7.5) are proportional to a monotonic function of the posteriors of any noise process, even in discrete time and asymptotically. However, if we measure the performance only with respect to the total input SNR, it will be seen that the performance is better than in the Gaussian case. Thus, the suboptimal class which requires only slight increase in the implementation complexity, is quite adequate.

We now turn to the asymptotic performance analysis. When $n \to \infty$, the $\{d_m\}$ are jointly Gaussian distributed; this can be shown for bounded $l(\cdot)$ by invoking Liapounoff's version of the central limit theorem (notice that the summands are functions of n, hence the simple CLT does not hold), see [64, p 201]. For linear $l(\cdot)$, the proof is elementary. Hence, computation of means and covariances is sufficient for the error probability. With tedious but straightforward computation it can be shown that:

$$\lim_{n \to \infty} E(n^{1/2} d_m \mid H_k) = A \left(\frac{2T}{N_0}\right)^{1/2} E_f(l') \epsilon \rho_{mk} \tag{7.6}$$

$$\lim_{n \to \infty} Cov(n^{1/2} d_m, n^{1/2} d_l \mid H_k) = E_f(l^2) \epsilon \rho_{ml} \tag{7.7}$$

$$\lim_{n \to \infty} Var(n^{1/2} d_m \mid H_k) = E_f(l^2) \epsilon \triangleq \sigma^2 \tag{7.8}$$

Since the signal amplitude is always $O(n^{-1/2})$, the above expressions are obtained by means of a Taylor series expansion. These expressions are similar to the corresponding ones for the matched-filter/Gaussian noise case, except for the $E(l')$ and $E(l^2)$ terms. Clearly, calculation of the error probabilities is unchanged if the decision variables are normalized to $\mathbf{N}(\mathbf{o}, [\rho_{ml}])$ according to

$$d_i \to v_i = \frac{n^{1/2} d_i - E(n^{1/2} d_i \mid H_k)}{\sigma}$$

Thus, the error probability is $P_e(k) = 1 - P_c(k)$, where

$$P_c(k) = Pr\{d_k = \max_i d_i \mid H_k\} = Pr\{\forall i, v_i \leq v_k + P^{1/2}(l, f)(1 - \rho_{ik}) \mid H_k\} \tag{7.9}$$

In that way, we get the same well known expressions for the error probability, for any signal set $\{s_m(t)\}$, but the signal to noise energy ratio (i.e., the factor ϵ/N_0 in Chapters 7-9 of [65]), is replaced by $P(l, f)$,

$$P(l, f) = \frac{2TA^2\epsilon}{N_0} G(l, f) \quad , \quad G(l, f) \triangleq \frac{E_f^2(l')}{E_f(l^2)} \tag{7.10}$$

We see that the error performance in the multiple hypothesis detection problem for the AWNGN channel is again characterized by the functional $G(l, f)$, and all the results

on robust estimation of a location parameter by an M-estimator apply here as well. From Eq.(7.9), it is seen that, for arbitrary signal set, the probability of correct decision is a monotonic increasing function of $P(l, f)$. To show this, denote $P^{1/2}=X$, and notice that if $X_2>X_1$, the event $A=\{\forall i, v_i \leq v_k+X_1(1-\rho_{ik}) \mid H_k, v_k\}$ is included in $B=\{\forall i, v_i \leq v_k+X_2(1-\rho_{ik}) \mid H_k, v_k\}$, since $1-\rho_{ik}>0$. Therefore $Pr\{A\}<Pr\{B\}$, and from (7.9) $P_c(X_1)=\int Pr\{A\}dF(v_k)<P_c(X_2)$. Consequently, whenever a minimax relation on $G(l, f)$ exist in some class $\{l \in \mathrm{L}\}$ and $\{f \in \mathrm{P}\}$, it is translated into a minimax relation within the above class of decision rules based on the limiter-correlators (7.5).

For the matched filters receiver, $l(x)=x$ and $G(l, f)=1/E_f(x^2)=1/C$, i.e., the inverse of the total noise power density. Thus, if one only measures the noise power at the signal bandwidth and does not have any knowledge of the shape of the noise distribution, the error probability will be exactly what one would have expected from the classical theory; hence, one would not have any reason to be alarmed. This might be an explanation why communication in a non-Gaussian noise has not received any attention so far. However, it is known that for all distributions of equal variance, the lowest $G(l, f)$ is achieved by the Gaussian, and the pair of the Gaussian distribution and a linear $l(\cdot)$ constitute a saddle-point, see [3, p. 83]. As a consequence of the monotonicity of the error probability with respect to the effective signal to noise ratio P, the performance of the matched filters receiver is a lower bound; any other receiver which utilizes some knowledge on the deviation from Gaussianity (within the postulated class of AWNGN processes) will perform better. In particular, if we assume that the marginals belong to a nominally Gaussian ϵ–mixture family, $f(x)=(1-\epsilon)g(x)+\epsilon h(x)$, the decision rule based on the limiter-correlator (7.5) with Huber's function $l_0(x; -K, K)$ of Eq. (4.9) is the minimax solution [5].

[5] When $k=0$, d_m correlates the signal with the sign of the observations. This scheme was considered by Beaulieu and others [72], in the binary case, for reasons of reducing the number of multiplications.

Let us write

$$P(l_0, f) = \frac{2TA^2\epsilon}{N_0 C} \cdot CG(l_0, f) \quad , \quad C = (1-\epsilon) + \epsilon\sigma_h^2$$

where σ_h^2 is the variance of the contaminating density. Thus, the first factor in $P(l_0, f)$ is the true signal to total noise energy ratio, while the second factor represents a gain factor which increases the effective SNR P for non-Gaussian noise. For fixed ϵ, $G(l_0, f)$ is bounded, but C can be arbitrary large when the contaminating noise is heavy tailed. Therefore, the performance improvement with respect to the matched filters receiver and Gaussian noise of equal power(C) can be arbitrary large! As an example, when h is also Gaussian and ϵ=0.1, gains of 1.4, 8.8 and 17.9 dB are obtained with σ_h =3, 10 and 30, respectively. ($G(l_0, f)$ are given by the inverse of the variances in Table 5.1).

The last conclusion can also be interpreted by an information-theoretic viewpoint: with the postulated AWNGN channel, essentially error-free communication can be achieved at transmission rates *much higher* than the channel capacity of the equivalent AWGN channel. Consider, for example, an orthogonal signal set - $\rho_{ij}=\delta_{ij}$. If the noise density f is known, the optimal performance for the assumed class of decision rules is obtained for $l(f)=-f'/f$, and then $G(l(f), f)=I(f)$, Fisher's information. The data transmission rate is $R=\log_2 M/T$ and $S=A^2\epsilon$ is the average signal power. Paralleling the derivation of [65, Chap. 8.4], we obtain that the limit of the error probability for large M is

$$\lim_{M\to\infty} P_e = \begin{cases} 1 & \text{if} \quad \frac{S}{N_0 R} I(f) < \ln 2 \\ 0 & \text{if} \quad \frac{S}{N_0 R} I(f) > \ln 2 \end{cases} \tag{7.11}$$

Since $S/(N_0 \ln 2)$ is the capacity of the infinite bandwidth AWGN channel, and since for all distributions of fixed variance $I(f)$ is minimized by the Gaussian, the previous claim is clear. [6] As a consequence of Shannon's coding theorem, this derivation shows, in fact,

[6] Indeed, theorem 7.4.3 of Gallager [66] states that for a given noise variance, and for constrained signal energy, Gaussian noise is the worst additive noise from a capacity viewpoint.

that the capacity of the postulated AWNGN channel is lower bounded by $I(f)\cdot S/(N_0 \ln 2)$ [7].

As a final comment, let us note that if the non-Gaussian noise is indeed a QMGP, identical results can be shown for the continuous time version of (7.5). The derivation parallels that of Martin [70], based on theorems of Rosenblatt [71] on the asymptotic normality of some integral functionals of non-Gaussian processes.

7.3 Robust M-ary Coherent receivers: Implementation with generalized M-estimators

If the channel is severely band limited, or if it is desired to reduce the number of computations in a digital implementation so that samples are taken at a lower rate, n might not be very large, and the input SNR needed for small error probabilities is bound to be relatively high. In this case the limiter-correlator structure might cause performance losses. Motivated by the very good small sample performance of M-estimators, and observing that d_k, when properly normalized by the energy, is an unbiased estimator of the signal, it seems appropriate to generate the decision variables as generalized M-estimators

$$d_k = \arg \{ \sum_{i=1}^{n} s_{ki}\, l(x_i - s_{ki}\, d_k) = 0 \} \qquad k = 1,.....M \tag{7.12}$$

Unfortunately, even the asymptotic performance is difficult to compute, for an arbitrary

[7] The direct strategy to compute capacity for an arbitrary additive channel was derived in [66]. If x is the channel's input and y its output, the constraint on the input energy is translated into a constraint on the output energy. Then, finding the capacity is reduced to maximizing the output entropy $H(Y)$, with respect to $P_Y(y)$, for the given constraint; afterwards, it is needed to find a distribution $P_X(x)$ yielding the maximizing $P_Y(y)$. Unfortunately, the maximizing $P_Y(y)$ is found to be Gaussian, therefore this approach yields results only for an additive Gaussian noise.

signal set. Indeed, when H_k is true, d_k is asymptotically Gaussian with

$$\lim_{n\to\infty} E_f\,(d_k \mid H_k) = a_n \epsilon \quad , \quad \lim_{n\to\infty} Var\,(d_k \mid H_k) = \frac{\epsilon}{2BT}\frac{E_f\,(l^2)}{E_f^2(l')}$$

This can be shown to be a special case of robust M-estimation of a one dimensional regression problem, for which asymptotic normality has been proven by Huber, see [3]. Moreover, it can be shown that each of the d_m, $m \neq k$ converge in probability to $a_n \epsilon \rho_{mk}$. However, it is difficult to show that the vector of decision variables is jointly normal with covariances corresponding to Eq. (7.7). Presumably, one might parallel the derivation following Eq. (D.11) of App. D. Furthermore, to reduce the complexity, a "one - step " implementation might be quite efficient, as has been shown in Chap. 5 for the simpler binary problem with a constant signal. The "one - step" decision variables are the first iteration of Newton's method, computed according to

$$d_k^{\,1} = d_k^{\,0} + \frac{\sum_{i=1}^{n} s_{ki}\, l\,(x_i - s_{ki}\, d_k^{\,0})}{\sum_{i=1}^{n} s_{ki}\, l'(x_i - s_{ki}\, d_k^{\,0})} \tag{7.13}$$

where $d_k^{\,0}$ is given by (7.5).

For several particular cases which are widely applied, it is possible to exhibit the asymptotic performance; these are the PAM, PSK and QAM signaling schemes.

Robust PAM

Here $s_{mi} = A_m\, s_i$, $A_m = 2m - 1 - M$. Of course, the energies are not equal. Now, as in the classical Gaussian case, only a single decision variable is needed

$$d \;= \arg\,\{\, \sum_{i=1}^{n} s_i\, l\,(x_i - d s_i) = 0 \,\} \tag{7.14}$$

and the decision is made in favor of one of the M possible amplitude levels $a_n A_m$ which is nearest to d (Of course, for PAM the channel attenuation on the transmitted signal must be precisely known). Clearly, the error performance is closely related to the

problem of estimating the one dimensional regression parameter $a_n A_m$ for the observations $x_i = a_n A_m s_i + n_i$. As $n \to \infty$, $n^{1/2}(d - a_n A_m)$ is zero mean Gaussian with variance $V = E_f(l^2)/\epsilon E_f^2(l')$. Hence, as in the Gaussian case (cf. [68]), the error probability is given by $P_e = 2((M-1)/M)\Phi(-2P(l, f)^{1/2})$. As before, monotonicity with respect to the generalized SNR leads to the previous conclusions on robustness.

Robust PSK and QAM

For PSK, the transmitted signal is

$$s_m(t) = \mathrm{Re}[U_m e^{j2\pi f_c t}] \quad , \quad U_m = e^{\frac{j2\pi(m-1)}{M}} \quad , \quad m = 1, \ldots M \tag{7.15}$$

As usual for PSK, we assume that the noise $w(t)$ is a narrowband bandpass process around the carrier f_c, i.e., $S_w(f) = CN_0/2$ for $|f \pm f_c| \leq B/2$ and zero otherwise, cf. [68]. We also assume $f_c > B/2 >> 1/2T$ so that $w(t)$ can be represented by its in-phase and quadrature low-pass processes. Following demodulation by the in-phase and quadrature mixers, amplitude normalization and time sampling, the analog to Eq. (7.3) is

$$H_m : x_{ni} = a_n Re(U_m) + n_{ci} \qquad y_{ni} = a_n Im(U_m) + n_{si} \quad i = 1, \ldots n = BT \tag{7.16}$$

With the previous assumptions, the marginal distributions of the $\{n_{ci}, n_{si}\}$ are identical, and all the r.v.'s are uncorrelated. We assume the analog for (p), which in this case requires also that if n_{ci} is uncorrelated with n_{si}, they are statistically independent. Actually, similar to the derivation of App. D, it suffices to require that the distribution of n_{ci} and n_{si} is circularly symmetric. Now, it is necessary to compute two M-estimators as decision variables

$$d_x = \arg\{ \sum_{i=1}^{n} l(x_i - d_x) = 0 \} \quad d_y = \arg\{ \sum_{i=1}^{n} l(y_i - d_y) = 0 \} \tag{7.17}$$

From them, an estimate of the phase is computed as $\hat{\theta} = \tan^{-1} d_y/d_x$, and a decision in favor of H_m is made if $\theta_m = 2\pi(m-1)/M$ is nearest to $\hat{\theta}$. d_x and d_y are asymptotically jointly Gaussian and independent, with common variance $E_f(l^2)/nE_f^2(l')$. Hence, the

expression for the error probability is identical to the Gaussian case where the SNR is replaced by $P(l, f)$, with ϵ=1. Consequently, the previous conclusions on robustness are valid. Moreover, one-step versions of (7.17) have identical asymptotic distribution; hence, the simplest implementation has the identical robustness property. Furthermore, since for PSK the equivalent lowpass signal is constant, identical results are obtained (with reduced computation load) if d_x and d_y are replaced by α–trimmed estimators, as in Chap. 5.2.

It is well known that the bandwidth efficiency of the PAM and PSK systems is penalized by the required high SNR, and that a combination of multiple signal amplitudes and phases in quadrature amplitude modulation (QAM) reduces the required SNR substantially. Since in QAM $s_m(t)=Re\,[(A_m+jB_m)\exp(j\,2\pi f_c\,t)]$, a receiver which decides in favor of the point (A_m, B_m) which has the shortest Euclidean distance to (d_x, d_y) of (7.17), will have the same robustness properties as the robust PSK receiver.

Orthogonal and equal-distance signals

Let us rewrite Eq. (7.3) as

$$x_{ni} = a_n \sum_{m=1}^{M} \theta_m s_{mi} + n_i \qquad i=1,....n \tag{7.18}$$

In this way, we can cast the problem as a regression problem, where we first estimate the vector of unknown parameters $\theta^t=(\theta_1,....,\theta_M)$ and then we decide H_k if $\hat{\theta_k}$ is the largest. The motivation for this modification is that we can utilize some results from the theory of robust regression; namely, the solution of the system

$$\sum_{i=1}^{n} s_{pi}\, l_0(x_i - \sum_{m=1}^{M} \hat{\theta}_m s_{mi}) = 0 \qquad p=1,.....,M \tag{7.19}$$

is the minimax optimal solution, see [3]. On the negative side, in general the coupled equations might require more computations than the solution of the M one dimensional problems (7.12). The second disadvantage is that we do not utilize the prior knowledge

on $\boldsymbol{\theta}$ (i.e., one of the components is 1 and the other are 0); this might result in performance losses. It turns out that for the two important cases of orthogonal and uniform-distance signal sets ($\rho_{ij}=\rho$ for $i\neq j$, $-1/(M-1)<\rho<1$), no losses are incurred. To show this let us define the S matrix whose m^{th} column is the signal vector on the m^{th} hypothesis $(s_{m1},\ldots..s_{mn})^t$, and $H=S(S^tS)^{-1}S^t$. From (7.4), $S^tS=n\epsilon R_d$, where R_d is the matrix with elements ρ_{ij}. Huber's condition for joint asymptotic normality of the solutions of (7.19) is that $h=\max_i H_{ii}\rightarrow 0$ as $n\rightarrow\infty$. For the uniform-distance set, it can be shown that R_d^{-1} has equal diagonal terms $((M-2)\rho+1)/D$, $D=1+\rho(M-2)-\rho^2(M-1)>0$, and equal off-diagonal terms $-\rho/D$. Thus, it can be shown that, with $G=\max_{i,j}s_{ij}$,

$$h \leq \max\{\frac{M(1+(M-1)\rho)G^2}{n\epsilon D}, \frac{M(1-\rho)G^2}{n\epsilon D}\}$$

This is also valid for the orthogonal set where $\rho=0$. Hence, if M is not too large, the condition is satisfied, the solutions of (7.19) converge in probability to the true values, and has the covariance matrix

$$R_{\hat{\theta}}=(S^tS)^{-1}\frac{E_f(l^2)}{E_f^2(l')}=R_d^{-1}\frac{1}{n\epsilon}\frac{E_f(l^2)}{E_f^2(l')} \tag{7.20}$$

From (7.6) to (7.8), it is seen that asymptotically, $R_d\hat{\theta}$ converges in probability to $\mathbf{d}$, where $\mathbf{d}$ is the vector of the limiter-correlators (7.5). Since the following equality converges in probability,

$$d_i-d_j=(1-\rho_{ij})(\theta_i-\theta_j)+\sum_{l\neq i\neq j}\theta_j(\rho_{il}-\rho_{jl})$$

we have for both the uniform-distance and orthogonal sets,

$$P_c(\mathbf{d}\mid H_k)=Pr\{\max_i(d_k-d_i)\geq 0\mid H_k\}=Pr\{\max_i(1-\rho)(\theta_k-\theta_i)\geq 0\mid H_k\}=P_c(\theta\mid H_k) \tag{7.21}$$

A simple counter-example shows that $\rho_{ij}=\rho$ is also necessary for (7.21). Note that the

above derivation is not valid for the *simplex* set, for which $\rho = -1/(M-1)$ and R_d is singular.

In conclusion, for most cases of interest, it has been shown that generating the decision variables from one form or another of M-estimation, leads to identical asymptotic error probability and robustness as of the limiter- correlator implementation. We conjecture superiority of the first for small sample sizes and small desired error probabilities; this, however must be confirmed by a simulation study, preferably with real non-Gaussian data.

7.4 Robust M-ary noncoherent communication

Similar results can be obtained for noncoherent communication, where no attempt is made to track the phase shift introduced by the channel θ. The only difference is that we have to consider the in-phase and quadrature components of the received signal. We demonstrate the structure for limiter- correlator implementation, but some of the derivations of Section 7.3 can also be extended.

Let the transmitted signal be $s_m(t) = Re\,[C_m(t)\exp(j\,2\pi f_c\,t)]$, $C_m(t) = A_m(t) - jB_m(t)$. The signal energy is given here by

$$\int_0^T s_m(t)dt = \frac{1}{2}\int_0^T |\,C_m(t)\,|^2 dt = \epsilon T \tag{7.22}$$

where we have assumed, as usual, $f_c\,T >> 1$ so that double frequency terms could be omitted. The following inner-products are defined

$$\mathrm{Re}\int_0^T C_m(t)C_l^*(t)dt = \int_0^T [A_m(t)A_l(t) + B_m(t)B_l(t)]dt = 2\epsilon T\,\hat{\rho}_{ml} \tag{7.23}$$

$$\mathrm{Im}\int_0^T C_m(t)C_l^*(t)dt = \int_0^T [A_m(t)B_l(t) - B_m(t)A_l(t)]dt = 2\epsilon T\,\tilde{\rho}_{ml} \tag{7.24}$$

where the same approximation leading to (7.22) has been used. The received signal under the m^{th} hypothesis is

$$r(t) = Re\,[(AC_m(t)e^{j\theta} + n_c(t) - jn_s(t))e^{j2\pi f_c t}] \quad 0 \leq t \leq T \tag{7.25}$$

where $n_c(t)$ and $n_s(t)$ are the in-phase and quadrature components of the narrowband bandpass noise process, with spectra $S_{n_c}(f) = S_{n_s}(f) = CN_0/2$ for $|f| \leq B/2$ and zero otherwise. For validity of the noise representation and for undistorted reception of the signal, the standard assumption $f_c > B/2 >> 1/T$ is made.

For the AWGN channel, the optimal receiver decides in favor of the largest d_m, where here d_m is the output of a quadrature matched filter, cf. [65]. The following representation of d_m, which can be shown to be identical to the more familiar expressions of [65], is more convenient for robust modification.

$$d_m = U_m^2 + V_m^2 \qquad U_m + jV_m = \int_0^T C_m(t)[X(t) + jY(t)]\,dt \tag{7.26}$$

where

$$X(t) = [2r(t)\cos(2\pi f_c t)]_{LPF} \qquad Y(t) = [2r(t)\sin(2\pi f_c t)]_{LPF} \tag{7.27}$$

and the symbol $[\cdot]_{LPF}$ denotes the operation of passing the argument through an ideal lowpass filter with unity gain and cutoff frequency $B/2$. Proceeding as in Section 7.2, the normalized in-phase and quadrature time samples on the m^{th} hypothesis are given by

$$H_m : \; x_{ni} = a_n\, Re\,[C_{mi}\, e^{-j\theta}] + n_{ci}\;,\; y_{ni} = a_n\, Im\,[C_{mi}\, e^{-j\theta}] + n_{si} \quad i = 1,..n = BT \tag{7.28}$$

The assumptions on $\{n_{ci}, n_{sj}\}$ are identical with those made following Eq. (7.16). As a robustification of the quadrature matched filter for the AWNGN channel, we introduce as before non-linear limiter-correlators. Namely, we still take $d_m = U_m^2 + V_m^2$ to get rid of the unknown phase, but now (7.26) is replaced by

$$U_m = \frac{1}{n}\sum_{i=1}^{n}[A_{mi}\,l(x_{ni}) + B_{mi}\,l(y_{ni})] \quad V_m = \frac{1}{n}\sum_{i=1}^{n}[-B_{mi}\,l(x_{ni}) + A_{mi}\,l(y_{ni})] \tag{7.29}$$

Following similar arguments as before, $\{U_m\,,\ V_m\}$ are asymptotically jointly Gaussian. After a tedious but straightforward Taylor expansion, the following moments are obtained :

$$\lim_{n\to\infty} E\,(n^{1/2}U_m \mid H_k) = 2A\,(\frac{2T}{N_0})^{1/2}\epsilon E_f\,(l')[\hat{\rho}_{mk}\cos\theta + \tilde{\rho}_{mk}\sin\theta] \tag{7.30}$$

$$\lim_{n\to\infty} E\,(n^{1/2}V_m \mid H_k) = 2A\,(\frac{2T}{N_0})^{1/2}\epsilon E_f\,(l')[\hat{\rho}_{mk}\sin\theta - \tilde{\rho}_{mk}\cos\theta] \tag{7.31}$$

$$\lim_{n\to\infty} Cov\,(n^{1/2}U_m\,,\ n^{1/2}U_l \mid H_k) = \lim_{n\to\infty} Cov\,(n^{1/2}V_m\,,\ n^{1/2}V_l \mid H_k) = 2\epsilon E_f\,(l^2)\hat{\rho}_{ml} \tag{7.32}$$

$$\lim_{n\to\infty} Cov\,(n^{1/2}U_m\,,\ n^{1/2}V_l \mid H_k) = 2\epsilon E_f\,(l^2)\tilde{\rho}_{ml} \tag{7.33}$$

The variances are obtained by substituting $\hat{\rho}_{mm}=1$. For Eqs. (7.32 -33), it suffices to assume that $\{n_{ci}\,,n_{ci}\}$ have a circularly symmetric distribution, or that they are independent. Comparing Eqs. (7.30 - 33) with the corresponding terms for the linear matched filters and Gaussian noise, cf. [65, Chaps. 7.2 and 8.10], we observe that they are identical, provided we multiply U_m and V_m by $(n\,/(N_0E_f\,(l^2)))^{1/2}$, and we modify the SNR to $P(l\,,\ f\,)$ according to (7.10). Hence, we arrive at the previous conclusions on robustness, provided the error probability of the optimal noncoherent receiver for the AWGN channel is a monotonic decreasing function of the SNR. While this expected behavior is apparent from all the performance curves given in the literature, it is not a simple matter to prove it from the various complicated error functions for noncoherent M-ary demodulation. (Unfortunately, the ordering of decision variables in the coherent case, that is evident in (7.9), is not preserved here. This is due to the presence of the unknown phase. Consequently, a general proof of monotonicity has not been found). An exception is the binary case with orthogonal signals, $\hat{\rho}_{ml}=\tilde{\rho}_{ml}=0$ (they are the optimal for noncoherent binary communication) for which the error probability is given by $P_e=(1/2)\exp(-\epsilon/(2N_0))$. The generalization for M-ary orthogonal signals, which are believed to be the optimal, is given in App. J.

Appendix A

Small Sample Performance of the Sign Detector

The sign detector (SD) is the simplest and probably the most popular non-parametric detector for a deterministic lowpass signal in noise with uncertain symmetrical density. Under several assumptions, it is known to be the UMP detector [32] or the minimax LORD detector [5] for those particular situations. Under the asymptotic weak-signal/large-sample ($n \to \infty, H_1 \to H_0$) assumptions, the ARE with respect to the linear detector (LD) at noise density $f(\cdot)$ is $ARE_{SD,\,LD} \approx 4(\sigma f(0))^2$. This is 0.637 for Gaussian noise, 2 for Laplace (double exponential) noise and in between for many other densities. Hence, the SD is generally considered to be very efficient for many situations of uncertainty.

Small sample analysis reveals that it may not perform as well, as the fine details of the actual small sample-size error probabilities are lost in the asymptotic analysis. The sign test is given by:

$$T(\mathbf{x}) = \sum_{i=1}^{n} u(x_i) \begin{cases} > t & , \quad H_1 \\ = t & , \quad H_1 \text{ with probability } c \\ < t & , \quad H_0 \end{cases} \tag{A.1}$$

where $u(\cdot)$ is the unity step function. Notice that a randomized test is required since $T(\mathbf{x})$ is a discrete binomially distributed r.v. Assuming symmetry of $f(\cdot)$, the false alarm probability is given by:

$$\alpha_0 = (\frac{1}{2})^n \begin{cases} \sum_{k=t+1}^{n} \begin{pmatrix} n \\ k \end{pmatrix} + c \begin{pmatrix} n \\ t \end{pmatrix} & , \quad t \leq n-1 \\ c & , \quad t = n \end{cases} \tag{A.2}$$

When n is small and the desired α_0 is very small, the second row of (A.2) must be chosen, and therefore the detection probability is also proportional to the randomization constant c which might be very small. Denoting $Prob_{H_1}\{x_i \geq 0\} = p$, it is given by

$$P_d = cp^n = \alpha_0 2^n p^n \qquad \leq \qquad \alpha_0 2^n \mid_{SNR \to \infty} \tag{A.3}$$

The critical values of the sample size such that for $n \leq n_c$ the detector becomes very poor ($P_d \leq 0.5$, *even* for $SNR \rightarrow \infty$) are found to be 12, 19 and 26 for $\alpha_0 = 10^{-4}, 10^{-6}, 10^{-8}$, respectively; these numbers are typical of many practical systems. A similar situation is encountered with the Wilcoxon detector, as its minimal false alarm probability (without randomization) is also 2^{-n}.

Appendix B

An Approximation for the Distribution of the Envelope Detector

In this appendix an approximation is derived for the distribution of the coherent envelope statistic: $R = I^2 + Q^2$, $I = (1/n) \sum_1^n x_i$, $Q = (1/n) \sum_1^n y_i$, where $\mathbf{x}$ and $\mathbf{y}$ are i.i.d. vectors which are also independent of each other. When the samples are Gaussian, R is one sided exponentially distributed. Thus, we are interested in an approximation which exhibits this exponential limiting behavior.

Formally, any univariate density $f(x)$ can be expanded in a series of orthonormal functions $\{\Phi_k(x)\}$ as

$$f(x) = \sum_{k=0}^{\infty} C_k \, \Phi_k(x) \tag{B.1}$$

which is L_2 convergent when all moments of f exist. A classical expansion in terms of the Gaussian density and its derivatives (which are proportional to Hermite polynomials) is the type A Gram-Charlier series [36], or its reordered version, the Edgeworth's form which, when applied to the distribution of the samples mean, gives the best order in powers of $n^{-1/2}$. Conditions for convergence of the infinite Edgeworth series are rather technical and implicit (they require conditions on f which is usually unknown); moreover, they are of little value from the statistical viewpoint where the important question is whether a *small* number of terms can furnish a reasonable approximation. In particular, while the Edgeworth expansion of the sample mean converges to the Gaussian density with finite terms as $n \to \infty$, when the expansion is carried out for finite n, adding more terms in the series above some optimal number usually results in a *worse fit*. See [36, Chap. 6] and [37, Chap. 16].

As pointed out in a footnote [36, Section 6.24], when an approximation for the dis-

tribution of a *positive* random variable is desired, it is plausible to use a family of functions which are supported only on $[0, \infty)$. In this way, the error from the tails of the Gaussian density and its derivatives at $x < 0$ would be eliminated. Since in our case the limiting one sided exponential distribution belongs to the Gamma family, the appropriate polynomials are those of Laguerre defined by [38]

$$L_i^a(x) = \frac{e^x x^{-a}}{i!} \frac{d^i}{dx^i}(e^{-x} x^{i+a}) \tag{B.2}$$

The orthonormality relation is

$$\int_0^\infty e^{-x} x^a L_i^a(x) L_j^a(x) = \frac{\Gamma(a+1+i)}{i!} \delta_{ij} \tag{B.3}$$

hence the expansion is

$$f(x) = \sum_{i=0}^{\infty} C_i \; e^{-x} x^a L_i^a(x) \tag{B.4}$$

where

$$C_i = \frac{i!}{\Gamma(a+i+1)} \int_0^\infty L_i^a(x) f(x) dx \tag{B.5}$$

The first four polynomials are

$$L_0^a(x) = 1 \tag{B.6}$$

$$L_1^a(x) = 1 + a - x \tag{B.7}$$

$$L_2^a(x) = \frac{1}{2}[(a+1)(a+2) - 2x(a+2) + x^2] \tag{B.8}$$

$$L_3^a(x) = \frac{1}{6}[(a+1)(a+2)(a+3) - 3x(a+2)(a+3) + 3x^2(a+3) - x^3] \tag{B.9}$$

Like the Edgeworth expansion, it is possible to eliminate the terms $i = 1, 2$ by a transformation. Defining $x = by$, the following expressions are obtained from (B.5) and (B.6)-(B.9):

$$C_0 = \frac{1}{b\,\Gamma(a+1)} \tag{B.10}$$

$$C_1 = \frac{1}{b\,\Gamma(a+2)}[(a+1) - \frac{\mu_1}{b}] \tag{B.11}$$

$$C_2 = \frac{1}{b\,\Gamma(a+3)}\left[(a+2)(a+1) - 2(a+2)\frac{\mu_1}{b} + \frac{\mu_2}{b^2}\right] \tag{B.12}$$

$$C_3 = \frac{1}{b\,\Gamma(a+4)}\left[(a+3)(a+2)(a+1) - 3(a+3)(a+2)\frac{\mu_1}{b^2} + 3(a+3)\frac{\mu_2}{b^2} - \frac{\mu_3}{b^3}\right] \tag{B.13}$$

where $\mu_i = E_x(x^i)$. Solving for $C_1 = C_2 = 0$, we obtain:

$$b = \frac{\mu_2 - \mu_1^2}{\mu_1} \triangleq \frac{1}{\beta} \tag{B.14}$$

$$a = \mu_1^2/(\mu_2 - \mu_1^2) - 1 \triangleq \alpha \tag{B.15}$$

Hence, upon substituting back with $y = \beta x$

$$f(x) = \frac{\beta^\alpha e^{-\beta x} x^{\alpha-1}}{\Gamma(\alpha)} + \frac{\beta^2}{\Gamma(\alpha+3)}\left[(\alpha+2)\mu_2 - \beta\mu_3\right]\beta^\alpha e^{-\beta x} x^{\alpha-1} L_3^{\alpha-1}(\beta x) + \cdots \tag{B.16}$$

The leading term is recognized as the Gamma density which, as $\alpha \to 1$, tends to the desired exponential density.

Integrating (B.16) to get the false alarm probability $P_r\{x > t\,\mu_1\} = P_r\{y > t\,\alpha\}$ is simplified by using the defining expression for the Laguerre polynomials (B.2). Some tedious algebra yields:

$$P_{fa} = P_r\{x > t\,\mu_1\} = [1 - I(\alpha, t\,\alpha)] + \\ + \frac{C_3'}{6}(\alpha, \mu_1, \mu_3) e^{-\alpha t}(\alpha t)^\alpha \left[2\alpha t(\alpha+2) - (\alpha t)^2 - (\alpha+1)(\alpha+2)\right] + \cdots \tag{B.17}$$

where

$$C_3' = \frac{C_3}{\beta} = \frac{\alpha}{\Gamma(\alpha+3)}\left[(\alpha+1)(\alpha+2) - \frac{\alpha^2\mu_3}{\mu_1^3}\right] \tag{B.18}$$

and

$$I(\alpha, \alpha t) \triangleq \int_0^{\alpha t} \frac{e^{-t} t^{\alpha-1} dt}{\Gamma(\alpha)} \tag{B.19}$$

is the incomplete Gamma function. It is clear from (B.17) that the second term vanishes for $t = 0$ as well as for $t = \infty$, thus fulfilling some of the properties of a proper distribution (since the first term is a proper distribution). However, it cannot be said whether or

not (B.16) can assume negative values, a common feature of all finite Gram-Charlier expansions. The following term in the expansion turns to be so complicated so it will not be given here.

To get specific results for the coherent envelope statistics R defined in the beginning of the appendix, we need to express the moments of R in terms of those of the i.i.d. samples, $E(R^m) = \frac{1}{n^m} E[(\sum_1^n x_i)^2 + (\sum_1^n y_i)^2]^m$. Using the moment generating function, tedious but straightforward manipulations yield

$$\mu_1 = E(R) = \frac{2}{n}\sigma_i^2 \tag{B.20}$$

$$\mu_2 = E(R^2) = \frac{2\sigma_i^4}{n^2}[4 + \frac{K}{n}] \quad , \quad K \triangleq \frac{m_4}{\sigma_i^4} - 3 \tag{B.21}$$

$$\mu_3 = E(R^3) = \frac{2\sigma_i^6}{n^3}[24 + \frac{18K}{n} + \frac{1}{n^2}(L - 15K)] \quad , \quad L \triangleq \frac{m_6}{\sigma_i^6} - 15 \tag{B.22}$$

where $\sigma_i^2 = E(x_i^2)$, $m_j = E(x_i^j)$ and K is the coefficient of kurtosis. When x_i are Gaussian, both K and L obviously vanish. Substituting these into (B.15) and (B.18) we obtain:

$$\alpha = (1 + K/2n)^{-1} \tag{B.23}$$

$$C_3' = \frac{\alpha}{\Gamma(\alpha + 3)}[3\alpha + 2 - \alpha^2(5 + \frac{9K}{2n} + \frac{L - 15K}{4n^2})] \tag{B.24}$$

Hence, the limiting behavior of the approximation with only the first four terms of (B.4) occurs when the number of samples n is much greater than the kurtosis ($K/n \to 0$). Then $\alpha \to 1$ and the first term in (B.17) approaches e^{-t} as required by the central limit theorem, though (unlike the Edgeworth expansion) the first term itself actually includes all powers of K/n in a highly nonlinear fashion. When $\alpha < 1$, a comparison with tables of the incomplete Gamma function shows that for sufficiently large t (such that the nominal false alarm probability $P_{fa} = e^{-t}$ is small), the outcome of Eq. (B.17) could be orders of magnitude higher than the P_{fa} as expected. In the second term of (B.17), all the expressions tend to a constant when $\alpha \to 1$ except C_3' which

vanishes identically when $\alpha = 1$. For other values of α, expansion in K/n gives

$$\lim_{\frac{K}{n} \to 0} C_3' = -\frac{K}{n} \frac{11}{12} + O(\frac{K^2}{n^2}) \tag{B.25}$$

For the normal-normal mixture: $f(x_i) = (1-\epsilon)\, N(0, 1) + \epsilon N(0, C^2)$,

$$K = 3\epsilon(1-\epsilon)\left[\frac{C^2-1}{1-\epsilon+\epsilon C^2}\right]^2 \quad \underset{\epsilon C^2 >> 1}{\longrightarrow} \frac{3}{\epsilon} \tag{B.26}$$

$$L = 15\left[\frac{(1-\epsilon+\epsilon C^6)}{(1-\epsilon+\epsilon C^2)^3} - 1\right] \quad \underset{\epsilon C^2 >> 1}{\longrightarrow} \frac{15}{\epsilon^2} \tag{B.27}$$

Therefore, with small amounts of large contamination and n not very large, α can be much smaller than one and the contribution of (B.24) is also not negligible.

Actual computation of (B.17), using Equations (B.20)-(B.24), (B.26), (B.27), is shown in figures 2.1-2.5 of Section 2.2 and compared to Monte-Carlo simulation results. For α roughly in the range [0.5,1], this approximation is seen to be quite good. It is still a reasonable prediction of the amount of increase in the false alarm probability down to $\alpha = 0.3$.

APPENDIX C

The Maximin Test for a Single Observation

For a single observation of a r.v. R which comes from simple mixture hypotheses $\mathbf{P}_0$ and $\mathbf{P}_1$ as in Eq. (1.1), Huber's [2] maximin robust test takes the form

$$d^*(R) = \begin{cases} H_1 & , \quad l(R\,;L'\,,L''\,) > t \\ H_1 \text{ with probability } c & , \quad l(R\,;L'\,,L''\,) = t \\ H_0 & , \quad l(R\,;L'\,,L''\,) < t \end{cases} \tag{C.1}$$

where the soft limited $l(R\,;L'\,,L''\,)$ is given by Eqs. (1.6-1.8). The constants $0 \leq L' \leq L'' < \infty$ are found by solving

$$P_1\{L \geq L'\} + L'\,P_0\{L < L'\} = (1-\epsilon)^{-1} \tag{C.2}$$

$$P_0\{L < L''\} + (1/L'')P_1\{L \geq L''\} = (1-\epsilon)^{-1} \tag{C.3}$$

where $L(R) = f_1(R)/f_0(R)$ is the LR under the nominal situation, and is assumed to be continuous. Application of these equations to the distribution of the coherent envelope of a Rayleigh narrowband signal in narrowband Gaussian noise, Eqs. (3.3-3.4), results in Eqs. (3.6-3.7).

Since $l(R\,;L'\,,L''\,)$ is a monotone increasing function of $L(R)$, which in turn is monotonic in R, Eq. (C.1) can be reformulated as a randomized test on R. The threshold and the randomization constant are functions of the least favorable p.d.f.'s. Three different cases are possible, according to $t = L'$, $L' < t < L''$ and $t = L''$, see Figure 1. Define R' and R'' by $L(R') = L'$, $L(R'') = L''$. Define also $\alpha(x) = \int_x^\infty f_0(r)dr$, $\beta(x) = \int_x^\infty f_1(r)dr$. Then,

$$\underline{\text{Case A}} \qquad t = L' \;\Rightarrow\; \epsilon < (\alpha_0 - \alpha(R'))/(1-\alpha(R')) \tag{C.4}$$

$$d^*(R) = \begin{cases} H_1 & , \quad R \geq R' \\ H_1 \text{ with probability } c & , \quad R < R' \end{cases} \tag{C.5}$$

where $0 \leq c \leq 1$ is given, as a function of the desired α_0 and of R' which solves (C.2), by

$$c = \frac{\alpha_0 - \epsilon - (1-\epsilon)\alpha(R')}{(1-\epsilon)\,(1-\alpha(R'))} \tag{C.6}$$

The detection probability under the least-favorable q_1^* is

$$\beta(d^*, q_1^*) = c + (1-\epsilon)(1-c)\,\beta(R') = 1 - (1-\alpha_0)L' \tag{C.7}$$

Case B $$L < t < L'' \Rightarrow \frac{\alpha_0 - \alpha(R')}{1-\alpha(R')} \leq \epsilon < \frac{\alpha_0 - \alpha(R'')}{1-\alpha(R'')} \tag{C.8}$$

$$d^*(R) = \begin{cases} H_1\,, & R \geq R_t \\ H_0\,, & R < R_t \end{cases} \tag{C.9}$$

where

$$\alpha_0 = \epsilon + (1-\epsilon)\,\alpha(R_t) \tag{C.10}$$

Here,

$$\beta(d^*, q_1^*) = (1-\epsilon)\,\beta(R_t) \tag{C.11}$$

Case C $$L'' = t \Rightarrow \epsilon \geq (\alpha_0 - \alpha(R''))/(1-\alpha(R'')) \tag{C.12}$$

$$d^*(R) = \begin{cases} H_1 \text{ with probability } c\,, & R \geq R'' \\ H_0\,, & R < R'' \end{cases} \tag{C.13}$$

where

$$c = \frac{\alpha_0}{\epsilon + (1-\epsilon)\,\alpha(R'')} \tag{C.14}$$

Here,

$$\beta(d^*, q_1^*) = \frac{\alpha_0(1-\epsilon)\,\beta(R'')}{\epsilon + (1-\epsilon)\,\alpha(R'')} = \alpha_0 L'' \tag{C.15}$$

Notice that in this case where large ϵ necessitates randomized maximin test, the resulting maximin lower bound on β, Eq. (C.15) is very small, as it is proportional to the desired α_0.

The proof of Eqs. (C.4-C.15) follows from (C.1-C.3). For Case A, (C.1) is translated into (C.5) as a result of the above mentioned monotonicity properties. To satisfy the constraint on the probability of false alarm, we must have

$$\alpha_0 = P_0^*\{R \geq R'\} + c\ P_0^*\{R < R'\}$$
$$= P_0^*\{R' \leq R < R''\} + P_0^*\{R \geq R''\} + cP_0^*\{R < R'\}$$

Utilizing the expression for the least favorable density q_0^*, Eq. (1.6),

$$\alpha_0 = (1-\epsilon)P_0\{R' \leq R \leq R''\} + c\,(1-\epsilon)P_0\{R < R'\} + \frac{(1-\epsilon)}{L''} P_1\{R \geq R''\}$$

The last term is equal to $1 - (1-\epsilon)P_0\{R < R''\}$, by virtue of Eq. (C.3). Collecting terms, we obtain

$$\alpha_0(1-\epsilon)(1-\alpha(R'))(c-1) + 1$$

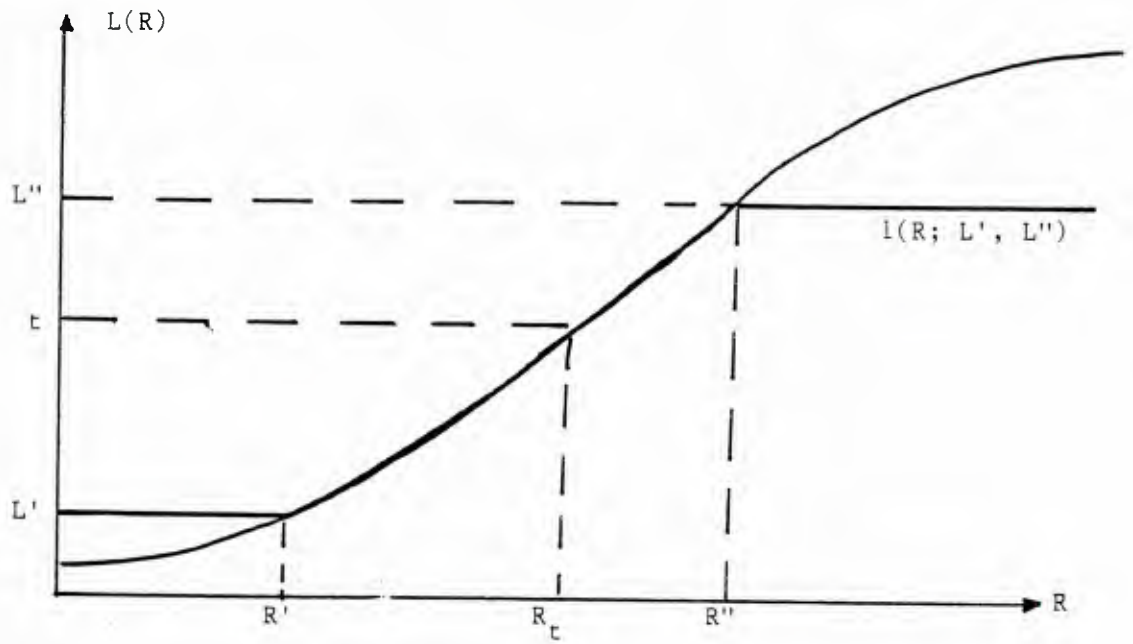
from which (C.6) follows. In order to satisfy $c \geq 0$, we get the appropriate ϵ range, Eq. (C.4). In a similar manner, from Eq. (1.2) for q_1^*,

$$\beta(d^*, q_1^*) = P_1^*\{R \geq R'\} + cP_1^*\{R < R'\}$$
$$= (1-\epsilon)P_1\{R \geq R'\} + (1-\epsilon)\,L'cP_0\{R < R'\}$$

and the last term, by virtue of Eq. (C.2), is equal to $c\,[1 - (1-\epsilon)P_1\{R \geq R'\}]$. Collecting terms, the first equality of Eq. (C.7) is obtained, and the second one is found by substitution of (C.6) and using (C.2) again.

Cases B and C are proved in a similar manner.

Figure 1



Appendix D

Statistics of the SSQME Test with Unmatched Frequency

We first generalize the asymptotic normality of M-estimators for a sequence of r.v.'s which are not identically distributed.

Lemma

Let $l \in \Psi$ of Section 4.1 and assume also that l is bounded. Let $\{x_i\}$ be a sequence of independent r.v.'s with p.d.f. $\frac{1}{\sigma_i} f\left(\frac{x_i - A m_i}{\sigma_i}\right)$ where f is symmetric and absolutely continuous, and $\lim_{n\to\infty} nA^2 = a^2 < \infty$. Then $\sqrt{n}\,\hat{A}_n$, where $\hat{A}_n = arg\left\{\sum_{i=1}^{n} l(x_i - \hat{A}_n) = 0\right\}$ is asymptotically distributed as $N(m, V)$ where

$$m = \frac{a \sum m_i E_i(l')}{\sum E_i(l')} \quad , \quad E_i(l') \triangleq \int l'(x) f\left(\frac{x}{\sigma_i}\right) \frac{dx}{\sigma_i} \tag{D.1}$$

and

$$V = \frac{\frac{1}{n}\sum E_i(l^2)}{\left[\frac{1}{n}\sum E_i(l')\right]^2} \quad , \quad E_i(l^2) \triangleq \int l^2(x) f\left(\frac{x}{\sigma_i}\right) \frac{dx}{\sigma_i} \tag{D.2}$$

Proof The proof follows Huber's [1] with some modifications to account for the unequal means and variances, hence it will only be briefly sketched. By monotonicity of $l(\cdot)$,

$$\lim_{n\to\infty} Prob\left\{\hat{A}_n \sqrt{n} \le k\right\} = \lim_{n\to\infty} Prob\left\{\sum l(x_i - k n^{-1/2}) \le 0\right\} \tag{D.3}$$

Define $y_{ni} = l(x_i - k n^{-1/2})$. Upon expanding in a Taylor series of powers of $n^{-1/2}$, and invoking the assumed symmetry properties of f and l, some terms vanish and we obtain:

$$m_{ni} \triangleq E(Y_{ni}) = n^{-1/2}(a m_i - k) E_i(l') + O(n^{-1}) \tag{D.4}$$

$$v_{ni} \triangleq Var(Y_{ni}) = E_i(l^2) + O(n^{-1}) \tag{D.5}$$

There is a slight complication since the y_{ni} are different for different values of n, and the usual formulations of normal convergence do not apply. However, this case is covered by the Lindeberg - Feller theorem [41], which asserts that a necessary condition for convergence to a standard normal distribution of $s_n = \sum_{i=1}^{r_n} (y_{ni} - m_{ni})/\sigma_n$ with $\sigma_n^2 = \sum_{i=1}^{r_n} v_{ni}$, is that each summand contributes negligibly to s_n in the sense of Lindeberg's condition. Explicitly,

$$\lim_{n \to \infty} \max_{1 \le i \le r_n} \frac{v_{ni}}{\sigma_n^2} = 0 \tag{D.6}$$

which is satisfied since $l(\cdot)$ is bounded, $v_{ni} \neq 0$ and σ_n^2 grows as n; The sufficient condition is proved in the same manner as in Huber[1,3]. Interchanging of $\lim_{n \to \infty}$ and $Prob\{\cdot\}$ is allowed since $\{\hat{A}_n\}$ converges. Subtracting $\lim_{n \to \infty} \sum m_{n_i}/\sigma_n = (m-k)/V^{1/2}$ from both sides of the argument of the second expression in (D.3) yields

$$\lim_{n \to \infty} Prob\{\hat{A}_n \sqrt{n} \le k\} = \Phi\left(\frac{k-m}{V^{1/2}}\right) \tag{D.7}$$

We now turn to prove Proposition 5.3. Let

$$\begin{bmatrix} I_i \\ Q_i \end{bmatrix} = \begin{bmatrix} c_i & s_i \\ -s_i & c_i \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} \tag{D.8}$$

where $c_i = \cos\theta_i$, $s_i = \sin\theta_i$, $\theta_i = 2\pi ik/n$. The input observations, with a sinusoid signal of Doppler frequency f_d, are given by

$$x_i = n_{c_i} + \frac{a}{\sqrt{n}} \cos(\phi + \phi_i) \ , \quad y_i = n_{s_i} + \frac{a}{\sqrt{n}} \sin(\phi + \phi_i) \tag{D.9}$$

where $\phi_i = 2\pi i f_d / f_r = 2\pi i \Delta / n$. (Δ measures the frequency deviation from the nearest filter, normalized by the filter bandwidth f_r/n). Let I and Q be the M - estimators of the $\{I_i\}$ and $\{Q_i\}$, respectively, where the $l(\cdot)$ function is assumed

monotonic, *bounded* and *skew – symmetric*. As a consequence of the symmetry of $l(\cdot)$ and the transformation (D.8), the following expectations under H_0 ($a=0$) satisfy:

$$E_0(l(I_i)) = E_0(l(Q_i)) = E_0(l''(I_i)) = E_0(l''(Q_i)) = 0$$

$$E_0(l') \triangleq E_0(l'(I_i)) = E_0(l'(Q_i)) \neq 0 \quad \text{(identical } \forall\, i\text{)}$$

$$E_0(l^2) \triangleq E_0(l^2(I_i)) = E_0(l^2(Q_i)) \neq 0 \quad \text{(identical } \forall\, i\text{)}$$

$$E_0(l(I_i)l(Q_i)) = E_0(l(I_i)l'(Q_i)) = E_0(l'(I_i)l(Q_i)) = 0$$

$$E_0(l'_I\, l'_Q) \triangleq E_0(l'(I_i)l'(Q_i)) \neq 0 \quad \text{(identical } \forall\, i\text{)}$$

Notice that I_i and Q_i are in general not independent under H_0 (unless X_i and Y_i are Gaussian), but they are uncorrelated, and their joint density is circularly symmetric - $f_0(I_i, Q_i) = g(I_i^2 + Q_i^2)$. Since the following derivation depends only on the above properties of the various expectations, which are valid with any bivariate circularly symmetric p.d.f., the original assumption that N_{c_i} and N_{s_i} are i.i.d. can be replaced by assuming they are circularly symmetric. In particular, this is true when the narrow - band noise process is realized in the "natural" way as $N_{c_i} = N_i \cos\gamma_i$, $N_{s_i} = N_i \sin\gamma_i$, where γ_i is $U[0, 2\pi]$ and is independent of N_i, which can be arbitrary positive r.v.

Applying the lemma, I and Q are marginally Gaussian. Since $Var_0(I_i) = Var_0(Q_i)$, (D.2) yields

$$\lim_{n\to\infty} Var_1(\sqrt{n}\, I) = \lim_{n\to\infty} Var_1(\sqrt{n}\, Q) = E_0(l^2)/E_0^2(l') \triangleq V_1 \qquad \text{(D.10)}$$

From (D.1),

$$\lim_{n\to\infty} E_1(\sqrt{n}\, I) = \lim_{n\to\infty} \frac{a}{n}\sum\cos(\phi+\phi_i) \triangleq \overline{I} \quad , \quad \lim_{n\to\infty} E_1(\sqrt{n}\, Q) = \lim_{n\to\infty} \frac{a}{n}\sum\sin(\phi+\phi_i) \triangleq \overline{Q} \qquad \text{(D.11)}$$

We want to show that the joint distribution of I and Q, averaged with respect to ϕ, is in the form of Eq.(4.15). For that, it is needed to show that I and Q are asymptotically independent. Invoking the monotonicity assumption,

(D.12)

$$P\{\sqrt{n}\,I \le c\,,\ \sqrt{n}\,Q \le d\} = P\{U_n \triangleq \frac{1}{\sqrt{n}}\sum U_{ni} \le 0,\ V_n \triangleq \frac{1}{\sqrt{n}}\sum V_{ni} \le 0\}$$

where

$$U_{ni} = l(I_i - \frac{c}{\sqrt{n}}) = l(N_{c_i}' + \frac{p_i}{\sqrt{n}})\ ,\ p_i = a\cos(\phi+\phi_i-\theta_i)-c \qquad \text{(D.13)}$$

$$V_{ni} = l(I_i - \frac{d}{\sqrt{n}}) = l(N_{s_i}' + \frac{q_i}{\sqrt{n}})\ ,\ q_i = a\sin(\phi+\phi_i-\theta_i)-d \qquad \text{(D.14)}$$

and the primed r.v.'s are the result of the transformation (D.8) applied to the input noise samples.

We have to show that the r.v. $W_n = \alpha U_n + \beta V_n = \sum_{i=1}^{n} W_{ni}$ is asymptotically Gaussian for any α and β. Let $Z_{ni} = (W_{ni} - E(W_{ni}))/(\sum_{i=1}^{n} Var(W_{ni}))^{1/2}$ and consider only bounded $l(\cdot)$ functions, which are sufficient for our study. (Actually, monotonicity is sufficient for the following, but the proof is longer). Obviously, $|Z_{ni}| \le M_{ni} < \infty$ and $\lim_{n\to\infty} \max_i M_{ni} = 0$ due to the above normalization. Thus, Lindeberg's condition is satisfied (see [64, pp. 201]), U_n and V_n (hence also I and Q) are asymptotically jointly Gaussian, and it suffices to show they are asymptotically uncorrelated for independence (Note that thus far no assumption on the joint distribution was necessary). Expanding in a Taylor series, noting that some of the various expectations under H_0 vanish as above, we find

$$E_1(U_{ni}) = \frac{p_i}{\sqrt{n}}E_0(l') + O(\frac{1}{n^{3/2}})\ ,\ E_1(V_{ni}) = \frac{q_i}{\sqrt{n}}E_0(l') + O(\frac{1}{n^{3/2}}) \qquad \text{(D.15)}$$

$$E_1(U_{ni}\,V_{ni}) = \frac{p_i\,q_i}{n}E_0(l'_I\,l'_Q) + O(\frac{1}{n^{3/2}}) \qquad \text{(D.16)}$$

$$Var_1(U_n) = Var_1(V_n) = E_0(l^2) + O(\frac{1}{n})$$

Hence,

$$E_1(U_n\,V_n) - E_1(U_n)E_1(V_n) = \frac{1}{n}\sum_{i=1}^{n}[E_1(U_{ni}\,V_{ni}) - E_1(U_{ni})E_1(V_{ni})]$$

$$= \left(\frac{1}{n^2}\sum_{i=1}^{n} p_i\,q_i\right)[E_0(l'_I\,l'_Q) - E_0^2(l')] + O(\frac{1}{n^{3/2}}) = O(\frac{1}{n})$$

Therefore, the asymptotic joint distribution of U_n and V_n is that of a pair of independent Gaussian variables. Returning to (D.12) and using (D.10 - D.11),

$$\lim_{n\to\infty} P\{\sqrt{n}\,I \le c\,,\ \sqrt{n}\,Q \le d\} = \lim_{n\to\infty} P\{U_n \le 0,\ V_n \le 0\}$$
$$= \lim_{n\to\infty} P\{\sqrt{n}\,I \le c\}P\{\sqrt{n}\,Q \le d\} = \Phi(\frac{c-\overline{I}}{V_1^{1/2}})\Phi(\frac{d-\overline{Q}}{V_1^{1/2}})$$

Transforming $\sqrt{n}\,I \to I$, $\sqrt{n}\,Q \to Q$, the argument of the exponent of the joint Gaussian density is thus

$$\frac{-1}{2V_1}\{\, I^2 + Q^2 + \overline{I}^2 + \overline{Q}^2 - 2(\overline{I}I + \overline{Q}Q)\,\}$$

The middle pair does not depend on ϕ

$$\frac{\overline{I}^2+\overline{Q}^2}{a^2} = \lim_{n\to\infty}[\frac{1}{n}\sum\cos(\phi+\phi_i)]^2 + [\frac{1}{n}\sum\sin(\phi+\phi_i)]^2 = \lim_{n\to\infty}\left|\frac{1}{n}\sum e^{j(\phi+\phi_i)}\right|^2 = \cdots$$
$$\cdots = \lim_{n\to\infty}[\frac{1}{n}\sum c_i]^2 + [\frac{1}{n}\sum s_i]^2 = \lim_{n\to\infty}\frac{\sin^2(n\pi f_d/f_r)}{n^2\sin^2(\pi f_d/f_r)} = \left[\frac{\sin(\pi\Delta)}{\pi\Delta}\right]^2 \triangleq G(\Delta)$$

where here $c_i \triangleq \cos(\phi_i)$ and $s_i \triangleq \sin(\phi_i)$, by straightforward calculation. Also,

$$\overline{I}I + \overline{Q}Q = \lim_{n\to\infty} a\sqrt{C^2+S^2}\cos(\phi - \tan^{-1}(\frac{S}{C}))$$

where $C \triangleq I(1/n)\sum c_i + Q(1/n)\sum s_i$ and $S \triangleq -I(1/n)\sum s_i + Q(1/n)\sum c_i$, from which $\lim_{n\to\infty} C^2+S^2 = (I^2+Q^2)G(\Delta)$ follows as can be verified by substituting for the definitions. We finally have for the asymptotic conditional joint density

(D.17)

$$f(I,Q \mid A,\phi) = \frac{1}{2\pi V_1}\exp\left[\frac{I^2+Q^2+nA^2G(\Delta)-2\sqrt{n}\,A\sqrt{(I^2+Q^2)G(\Delta)}\cos(\phi-\tan^{-1}(\frac{S}{C}))}{-2V_1}\right]$$

Averaging this with respect to ϕ on $[0,2\pi]$ gives the same expression as in section 4.2, since S and C do not depend on ϕ, and where the effective SNR is replaced by $nA^2G(\Delta)/2V_1$. Hence, the detection probabilities for the various SSQME tests of Section 5.2 are identical with the case $f_d = 0$ if the effective SNR is attenuated by $G(\Delta)$.

In the FFT processor, the sidelobes can be reduced, and the response can be flattened over the mainlobes of the individual filters ($| nf_d/f_r | \leq 0.5$), at a price of reduced SNR at $f_d = 0$, by appropriate weighting (e.g., Hamming's): $\{ \mathbf{x},\mathbf{y}\} \rightarrow \{\mathbf{w}^T\mathbf{x},\mathbf{w}^T\mathbf{y}\}$, which changes $G(\Delta) \rightarrow | \frac{1}{n} \sum_{k=1}^{n} w_k \exp(j 2\pi k \Delta/n) |^2$. The same can be done with the SSQME test. The derivation of the test statistic is similar to that above, but now we have to account for the unequal variances as a consequence of the weighting. The frequency response will be given by $G(\Delta) = | \frac{1}{n} \sum_k w_k' \exp(j 2\pi k \Delta/n) |^2$, where $w_k' \triangleq E_k(l')/\sum_k E_k(l')$. The design of the weighting coefficients is more complicated than in the linear FFT case; first $\{w_k'\}$ are chosen for the desired response, and then $\int l'(w_k x) f(x) dx = w_k'$ has to be solved for $\{w_k\}$.

Appendix E

The Efficacy of Quadrature Tests

Pitman's efficacy is usually employed as an asymptotic performance measure of test statistics, whose asymptotic distribution is Gaussian, c.f. [21], [25]. However, the asymptotic distribution of the quadrature type tests, which are studied in this thesis for detection of coherent fading narrowband signals, is non-Gaussian. Hence, an appropriate definition of efficacy is needed.

Let $Q_n = T_n^2(\mathbf{x}) + T_n^2(\mathbf{y})$, where the statistics $T_n(\cdot)$ are asymptotically Gaussian, under *both* the hypothesis and a location shift alternative: $f_1(x_i) = f_0(x_i - a)$, with

$$E_0 T_n(\cdot) = 0, \quad V_0 T_n(\cdot) = \sigma^2 \tag{E.1}$$

$$E_1 T_n(\cdot) = \sqrt{n}\, g(a), \quad V_1 T_n(\cdot) = \sigma^2 + O(a) \tag{E.2}$$

Assume that $g(0) = 0$, $0 < |g'(0)| < \infty$. For narrowband slow fading signals,

$$f(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^{n} f_0(x_i - aA\cos\theta,\, y_i - aA\sin\theta) \tag{E.3}$$

where A is a positive r.v. with arbitrary distribution, and θ is uniform r.v. on $[0,2\pi]$. Notice that, with proper normalization, A is a dimensionless r.v., a^2 represents the SNR, and the influence of the test statistic $T_n(\cdot)$ is gauged by $g(\cdot)$ and σ. Define

$$m_x \triangleq E_1 T_n(\mathbf{x}) = \sqrt{n}\, g(aA\cos\theta),\; m_y \triangleq E_1 T_n(\mathbf{y}) = \sqrt{n}\, g(aA\sin\theta) \tag{E.4}$$

and assume also that $T_n(\mathbf{x})$ and $T_n(\mathbf{y})$ are jointly Gaussian and

$$\lim_{n\to\infty} E_1\{(T_n(\mathbf{x}) - m_x)(T_n(\mathbf{y}) - m_y)\} = O(a)$$

i.e., $T_n(\mathbf{x})$ and $T_n(\mathbf{y})$ are asymptotically uncorrelated when $a \to 0$.

Definition The efficacy of the test Q_n is

$$\epsilon_{Q_n} \triangleq \frac{[g'(0)]^2}{\sigma^2} = \epsilon_{T_n} = \lim_{n\to\infty} \frac{1}{n}\left[\frac{dE_1 T_n / da \mid_{a=0}}{V_0^{1/2} T_n}\right]^2 \tag{E.5}$$

which is recognized as the common definition of the efficacy of T_n.

Justification We have to show that ϵ_{Q_n} parametrizes the asymptotic power function.

For that, the conditional (on A and θ) asymptotic distribution of Q_n is needed. To avoid singular detection, consider a sequence of alternatives $a_n = \delta n^{-1/2}$. By virtue of our assumptions on the asymptotic distributions of $T_n(\mathbf{x})$ and $T_n(\mathbf{y})$, their asymptotic distribution is bivariate Gaussian with zero correlation. Therefore, the asymptotic conditional p.d.f. is non-central Chi-squared with two degrees of freedom and non-centrality parameter $\lambda = m_x^2 + m_y^2$, c.f. [51, pp. 113-118],

$$f_{1\infty}(Q_n \mid A, \theta) \triangleq \lim_{n \to \infty} \frac{1}{2\sigma^2} \exp\left\{-\frac{Q_n + nh(a_n)}{2\sigma^2}\right\} I_0\left[\frac{\sqrt{Q_n\, nh(a_n)}}{\sigma^2}\right] \tag{E.6}$$

where

$$h(a) = g^2(aA\cos\theta) + g^2(aA\sin\theta) \tag{E.7}$$

Under H_0, Eq. (E.6) reduces to a one-sided exponential, thus $\alpha = P_0\{Q_n > t\} = \exp(-t/2\sigma^2)$. Under H_1,

$$\beta = \lim_{n \to \infty} E_A E_\theta P_r\{Q_n > -2\sigma^2 \log\alpha\} = E_A E_\theta \int_{-2\sigma^2\log\alpha}^{\infty} f_{1\infty}(Q_n \mid A, \theta) dQ_n \tag{E.8}$$

where interchanging of the limit and integrations is justified by the dominant convergence theorem. Expanding Eq.(E.7) in a Taylor series and utilizing $g(0)=0$, $|g'(0)| < \infty$, the first two terms vanish and we obtain

$$\lim_{n \to \infty} nh(a_n) = [\delta A g'(0)]^2 + O(1/\sqrt{n}) \tag{E.9}$$

which is not a function of θ, to order $n^{-1/2}$. Thus, with the definition of Marcum's Q-function, the limiting power is

$$\beta = E_A\, Q[\delta A\, |g'(0)|/\sigma,\ \sqrt{-2\log\alpha}] \tag{E.10}$$

For a given distribution of A, and a given input SNR (δ), the detectability is thus governed by the parameter $|g'(0)|/\sigma$, and the definition is justified. (Recall, however, that the random phase and amplitude incur detectability loss in terms of the required input SNR, c.f. [13], [51].)

Appendix F

Efficacy Optimal Tests for "Single Sweep" Ranking

Let $\{x_{i0}\}_{i=1}^{n}$ be the observables which contain the information on the alternative K, and $\{x_{ik}\}_{i=1}^{n}$, $k = 1,...,M$ be noise reference observations. Under K, a location shift model is assumed

$$f(\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_M) = \prod_{i=1}^{n} f(x_{i0} - A) \prod_{k=1}^{M} f(x_{ik}) \tag{F.1}$$

In radar terminology, during each "sweep" i (the repetition interval of the transmitted pulse train), M reference observations are taken in a region around the possible signal location.

A "single sweep" rank statistic is defined as

$$T_n = \sum_{i=1}^{n} a_{M+1}(R_{M+1,i}) \tag{F.2}$$

where $R_{M+1,i}$ is the rank of x_{i0} in $\{x_{ik}\}_{k=0}^{M}$. In the following, the $M+1$ subscript will be dropped for convenience.

As T_n is the sum of i.i.d. random variables, it is asymptotically Gaussian with

$$E\{T_n\} = nE\{a(R)\},\ Var\{T_n\} = nVar\{a(R)\} \tag{F.3}$$

The discrete probability distributions of the ranks is given by

$$P_1(l) \triangleq P_r\{R_i = l \in [1,...,M+1] \mid K\} = P_r\{x_{i(l-1)} \leq x_{i0} < x_{i(l)}\} \tag{F.4}$$
$$= \binom{M}{l-1} \int_{-\infty}^{\infty} f(x - A) F^{l-1}(x)[1 - F(x)]^{M+1-l}\, dx$$

and

$$P_0(l) \triangleq P_r\{R_i = l \mid H\} = 1/(M+1) \tag{F.5}$$

Thus,

$$E\{a(R) \mid K\} = \sum_{l=1}^{M+1} a(l) P_1(l),\ E_0(a) = \frac{1}{M+1} \sum_{l=1}^{M+1} a(l) \tag{F.6}$$

and

$$\frac{\partial E\{a(R)\mid K\}}{\partial A}\Big|_{A=0}=\frac{1}{M+1}\sum_{l=1}^{M+1}a(l)a(l,f) \tag{F.7}$$

where

$$a(l,f)\triangleq -(M+1)\binom{M}{l-1}\int_{-\infty}^{\infty}f'(x)F^{l-1}(x)[1-F(x)]^{M+1-l}\,dx \tag{F.8}$$

Also,

$$Var\{a(R)\mid H\}=\sum_{l=1}^{M+1}P_0(l)[a(l)-E_0(a)]^2=\frac{1}{M+1}\sum_{l=1}^{M+1}[a(l)-E_0(a)]^2 \tag{F.9}$$

and it can be easily verified that

$$\lim_{K\to H} Var\{a(R)\mid K\}=Var\{a(R)\mid H\}$$

Pitman's conditions [25] are satisfied, and the efficacy is given by

$$\epsilon=\frac{\left[\partial E\{a(R)\mid K\}/\partial A\Big|_{A=0}\right]^2}{Var\{a(R)\mid H\}}=\frac{\frac{1}{M+1}\left[\sum_{l=1}^{M+1}(a(l)-E_0(a))\,a(l,f)\right]^2}{\sum_{l=1}^{M+1}(a(l)-E_0(a))^2} \tag{F.10}$$

where we have used $\sum_{1}^{M+1}a(l,f)=0$, as can be shown from its definition, Eq. (F.8). Eq. (F.10) can be used to measure the efficacy of an arbitrary score function $a(\cdot)$ at a given noise p.d.f. By the Cauchy-Schwarz inequality, ϵ is maximized by $a(l)-E_0(a)=a(l,f)\Rightarrow a(l)=a(l,f)$, and

$$\max_{\{a(l)\}}\epsilon=\frac{1}{M+1}\sum_{l=1}^{M+1}a^2(l,f) \tag{F.11}$$

when the signal is time varying, $A\to s_iA$ in (F.1). Due to the sweep-to-sweep independence of the R_i, the influence of the s_i on ϵ appears as a multiplying factor. Following the same steps, it is readily shown that the optimal test is

$$T_n=\sum_{i=1}^{n}s_i\,a_{M+1}(R_{M+1,i}) \tag{F.12}$$

and Eq. (F.11) should be multiplied by $\sum_{i=1}^{n} s_i^2$.

For large M, the optimal scores can be replaced by the approximation [49] $a(l,f) \rightarrow \phi(\frac{l}{M+2}, f)$, where $\phi(u,f) = -f'(F^{-1}(u))/f(F^{-1}(u))$, and then $\max \epsilon \rightarrow I(f)$. Hence, when M is large, the "single sweep" ranking structure does not entail any losses.

It is easy to show that T_n of Eq. (F.2) with the optimal scores $a(l,f)$ is also locally optimal for finite n, since $T_{loc} = \sum_{i=1}^{n} P_1'(R_i)/P_0(R_i)$.

The $a(l,f)$ are also the optimal scores of the corresponding quadrature test $[\sum_{1}^{n} a(R_i^x)]^2 + [\sum_{1}^{n} a(R_i^y)]^2$, since its efficacy is identical with (F.10), see Appendix E.

Appendix G

The ARE of two-sample and single-sweep rank tests with Wilcoxon scores

From Eq.(F.10) of App.F, and from Eq.(6.3.12)

$$ARE_{2S,SS} = (1-\frac{\overline{S}^2}{M+1})\cdot\frac{[\int_{-\infty}^{\infty} J'(F(x))f^2(x)dx]^2}{\int_0^1 J^2(u)du}\cdot\frac{\sum_{i=1}^{N} a^2(i)}{\frac{1}{N}[\sum_{i=1}^{N} a(i)a(i,f)]^2} \qquad (G.1)$$

where $N=M+1$, $a(i,f)$ is given by Eq.(F.8), and the indices 2S and SS stand for two-sample and single-sweep test, respectively.

With Wilcoxon scores, $J(u)=u-1/2$, and $a(i)=i/(N+1)-1/2$. The ratio of integrals in (G.1) is easily computed to be $12[\int f^2(x)dx]^2$. Also, $\sum_{i=1}^{N} a^2(i)=(N-1)N/(12(N+1))$. For the remaining term,

$$\begin{aligned} A &\triangleq \sum_{i=1}^{N}(\frac{i}{N+1}-\frac{1}{2})a(i,f) = \frac{1}{N+1}\sum_{i=1}^{N} ia(i,f)= \\ &= -\frac{N}{N+1}\sum_{i=1}^{N} i\binom{N-1}{i-1}\int f'(x)F^{i-1}(x)[1-F(x)]^{N-i}\,dx \\ &= -\frac{N}{N+1}\sum_{k=0}^{N-1}(1+k)\binom{N-1}{k}\int f'F^k(x)[1-F(x)]^{N-1-k}\,dx \end{aligned}$$

Writing this as the sum of two terms by breaking the $(1+k)$, the term corresponding to the 1 vanishes after changing the order of summation and integration and utilizing the binomial expansion,

$$\sum_{k=0}^{N-1}\binom{N-1}{k}F^k[1-F]^{N-1-k} = 1$$

since $\int_{-\infty}^{\infty} f'(x)dx=0$. For the term corresponding to k, the $k=0$ term in the summation vanishes, thus another change of variables yields

$$A = -\frac{N(N-1)}{N+1}\sum_{l=0}^{N-2}\binom{N-2}{l}\int f'(x)F^{l+1}(x)[1-F(x)]^{N-2-l}\,dx$$

Interchanging the order of summation and integration and using the binomial expansion again, and integrating by parts, yields

$$A = -\frac{N(N-1)}{N+1}\int_{-\infty}^{\infty} f'(x)F(x)dx = \frac{N(N-1)}{N+1}\int_{-\infty}^{\infty} f^2(x)dx$$

Combining all the above derivations, (G.1) reduces to

$$ARE_{2S,SS} = (1+\frac{2}{M})(1-\frac{\overline{S}^2}{M+1}) \qquad \text{(G.2)}$$

which is independent of $f(\cdot)$.

Appendix H

Efficacy and Integration Gain for Non-coherent Stochastic Signals

Let the available observations under the alternative be the in-phase and quadrature components of a narrowband process $\{x_i, y_i\}_{i=1}^n$. The densities are given by

$$f_1(\mathbf{x},\mathbf{y};a) = \int\cdots\int f(x_1 - as_{c1}, y_1 - as_{s1}, \cdots, x_n - as_{cn}, y_n - as_{sn}) \cdot dF_s(s_{c1}, s_{s1}, \ldots, s_{cn}, s_{sn})$$

and $f_0(\mathbf{x},\mathbf{y}) = f_1(\mathbf{x},\mathbf{y};0)$. Here $f(\cdot)$ is the joint density of the narrowband noise, $F_s(\cdot)$ is the joint distribution of the additive stochastic signal, and the parameter a is related to the SNR. The following assumptions are made: 1) all first and second derivatives of $f(\cdot)$ exist; $\partial f / \partial x_k \partial y_l$ is denoted $f_{x_k y_l}$, and so on. 2) $\{\mathbf{s}_c, \mathbf{s}_s\}$ is a wide sense stationary process, with $Es_{ci} = Es_{si} = 0$. The absence of a common mean can actually be viewed as a definition of a non-coherent stochastic signal. In particular, it is true when $s_{ci} = A_i \cos\theta_i$, $s_{si} = A_i \sin\theta_i$, and the phases θ_i are i.i.d.

Consider a memoryless test statistic

$$T = \sum_{i=1}^{n} g(x_i, y_i)$$

Under rather mild regularity conditions, T is asymptotically normal, and the asymptotic power is

$$\beta = \lim_{\substack{n \to \infty \\ a \to 0 \\ an^{\delta} \to c}} \phi\left(\frac{E_1 T - E_0 T}{V_1^{1/2} T} - \frac{V_0^{1/2} T}{V_1^{1/2} T} \phi^{-1}(1-\alpha) \right) = \phi(d - \phi^{-1}(1-\alpha))$$

for some $\delta > 0$, where

$$d = \lim_{\substack{n \to \infty \\ a \to 0 \\ an^{\delta} \to c}} \frac{E_1 T - E_0 T}{V_0^{1/2} T} = \epsilon^{1/2}$$

Expanding the numerator in a Taylor series

$$E_1 T - E_0 T = a \frac{dE_1 T}{da} + \frac{a^2}{2} \frac{d^2 E_1 T}{da^2} + \cdots$$

but

$$\frac{dE_1T}{da} = -E\left\{\left[\sum_{i=1}^{n}(s_{ci}\,f_{x_i}+s_{si}\,f_{y_i})\right]\left[\sum_{j=1}^{n}g(x_j\,,\,y_j)\right]\right\} = 0$$

where the expectation is taken over all the random variables. Thus, unlike the coherent signal cases, the first term in d is proportional to a^2 rather than to a, under the most general assumptions. This also implies that a non-trivial definition of a locally optimal detector for non-coherent signals, is the one that maximizes the *second* derivative of the power function at zero.

To get a simple expression for the efficacy, we assume now that the noise samples are i.i.d., and that the signal is uncorrelated and its spectrum is symmetric around the carrier frequency. These, together with the w.s.s. assumption, amount to (c.f. [14], [51])

$$Es_{ci}^2 = Es_{si}^2 = 1,\;\; Es_{ci}\,s_{cj} = Es_{si}\,s_{sj} = 0,\;\; Es_{ci}\,s_{sj} = 0$$

Thus, it is readily found that

$$V_0T = nV_0g\,,\; E_0T = nE_0g$$

$$\frac{d^2E_1T}{da^2}\bigg|_{a=0} = n\int\int g(x\,,y)[f_{xx}(x\,,y)+f_{yy}(x\,,y)]dxdy \triangleq nI(g\,,f\,)$$

Hence,

$$d = \frac{\sqrt{n}\,a^2}{2}\,\frac{I(g\,,f\,)}{V_0^{1/2}g}$$

By way of comparison, for coherent signals d is proportional to a, and is larger for weak signals. If the narrowband noise is normalized to unity variance, $a^2/2 = SNR$. The *integration gain* for two detectors with the same $g(\cdot)$, is defined as the limit of the ratio of the input SNRs, when both operate at the same level. Thus,

$$IG_{1,2} \triangleq \lim_{\substack{n_1,n_2\to\infty \\ \alpha_1=\alpha_2=\alpha \\ \beta_1=\beta_2\neq\alpha}} \frac{SNR_2}{SNR_1} = \sqrt{\frac{n_1}{n_2}}$$

By way of comparison, the integration gain of detectors for coherent signals is (n_1/n_2), therefore coherent signals lead to better detection performance.

Appendix I

The efficacy and asymptotic distribution of rank Doppler tests

I.1) One - and two - sample rank Doppler tests

With the definitions of Chap. 6.6.1, the means of the observations after the "butterfly" transformations, Eq.(6.6.2), are

$$d_i^I \triangleq E_1 I_i = aAs_i \cos\delta_i \,[1-U(i-n-1)] \quad , \quad \delta_i \triangleq 2\pi i\,\Delta/n + \theta \tag{I.1}$$

$$d_i^Q \triangleq E_1 Q_i = aAs_i \sin\delta_i \,[1-U(i-n-1)] \tag{I.2}$$

Throughout this appendix, all expectations and variances under H_1 are conditional expectations, given A and θ. To compute the efficacy, the asymptotic means and variances of I and Q, defined in (6.6.4), are needed. They can be found from Theorem 6.2.9, provided the conditions (6.2.23) and (6.2.24) are satisfied. For the I channel, from (I.1),

$$A_n = \sum_{i=1}^{N} (d_i^I - \overline{d_i^I})^2 = na^2A^2(\overline{C^2} - \frac{\overline{C}^2}{M+1}) \leq na^2A^2\overline{C^2}$$

where $\overline{C} \triangleq (1/n)\sum_{i=1}^{n} c_i$, $\overline{C^2} \triangleq (1/n)\sum_{i=1}^{n} c_i^2$, $c_i \triangleq s_i \cos\delta_i$. Thus,

$$na^2A^2\overline{C^2} \leq a^2A^2\sum_{i=1}^{n} s_i^2 = na^2A^2\overline{S^2} < \infty$$

Hence, for $na^2 \rightarrow$ constant and $\overline{S^2} < \infty$, $\lim_{n\to\infty} A_n < \infty$ and (6.2.23) is satisfied. The condition (6.2.24) can be written with the above definitions as

$$B_n = \frac{n\,[\overline{C^2} - \overline{C}^2/(M+1)]}{\max_i \{(c_i - \overline{C}/(M+1))^2\,,\,(\overline{C}/(M+1))^2\}} \tag{I.3}$$

For an absolutely summable signal sequence, $(1/n)\sum_{i=1}^{n} |\,s_i\,| < \infty$, all the expressions in (I.3) are finite, thus $\lim_{n\to\infty} B_n = \infty$ as required. The conditions for the Q channel are satisfied in a similar manner.

Invoking Th. 6.2.9, we have

$$\mu_I \triangleq E_1 I = \frac{1}{\sqrt{n}} aA \sum_{i=1}^{n} (s_i - \frac{\overline{S}}{M+1}) s_i \cos\delta_i \; G'(J\,,\,f\,) \tag{I.4}$$

$$\mu_Q \triangleq E_1 Q = \frac{1}{\sqrt{n}} aA \sum_{i=1}^{n} (s_i - \frac{\overline{S}}{M+1}) s_i \sin\delta_i \, G'(J, f) \tag{I.5}$$

where $G'(J, f)$ is the integral in (6.2.27). From (6.2.28) and (6.3.14),

$$V_0 I = V_0 Q = \left(1 - \frac{\overline{S}^2}{M+1}\right) \int_o^1 J^2(u) du \tag{I.6}$$

As in App. D, the derivation following Eq.(D.7), θ cancels when summing $\mu_I^2 + \mu_Q^2$. Hence, the efficacy

$$\epsilon_{T_n} = \lim_{n \to \infty} \frac{(d\mu_I / da)^2 + (d\mu_Q / da)^2}{V_0 I} \tag{I.7}$$

is found when substituting (6.6.5) into (6.3.11)

I.2) Single - sweep rank Doppler test

$T_n(k)$ of Eq. (6.6.11) can be written as $T_n(k) = I^2 + Q^2$, where here

$$I = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_i = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (u_i c_i + v_i s_i), \quad Q = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Q_i = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (-u_i s_i + v_i c_i) \tag{I.8}$$

$\{u_i, v_i\}$ are the scored single - sweep ranks defined in (6.6.10), and here

$$c_i \triangleq \cos\alpha_i, \; s_i \triangleq \sin\alpha_i, \; \alpha_i \triangleq 2\pi ik/n \tag{I.9}$$

From the original assumptions, $\{U_i\}$ and $\{V_i\}$ are sequences of independent r.v., and U_i is independent of V_j. Though $\{I_i\}$ and $\{Q_i\}$ are not identically distributed, Liapounoff's version of the central limit theorem is valid for I and Q, since the third centered absolute moments of I_i and Q_i are bounded, as they are discrete and bounded r.v. Paralleling App. D, to prove (6.6.12) we have to show the following:

a) $$\lim_{H_1 \to H_0} \frac{E_1^2(I)}{Var_1(I)} + \frac{E_1^2(Q)}{Var_1(Q)} \sim G(\Delta) = [\sin(\pi\Delta)/\pi\Delta]^2$$

b) $$\lim_{H_1 \to H_0} \frac{IE_1(I)}{Var_1(I)} + \frac{QE_1(Q)}{Var_1(Q)} \sim \sqrt{I^2+Q^2}\cos(\theta - \tan^{-1}\delta(I, Q))$$

c) $$\lim_{H_1 \to H_0} \frac{E_1(IQ) - E_1(I)E_1(Q)}{\sqrt{Var_1(I)Var_1(Q)}} = 0$$

Here, $\lim_{H_1 \to H_0}$ is a shortened notation for the limit ($n \to \infty$, $a \to 0$, $na^2 < \infty$), $\sim$ denotes

proportional to, and $\delta(I, Q)$ is some function $R^2 \rightarrow R^1$.

Throughout, we assume the score function $a_{M+1}(l)$ is skew - symmetric, thus $E_0(U_i)=E_0(V_i)=E_0(I)=E_0(Q)=0$. The observations model is of Eq.(6.3.1), where $\theta \rightarrow \beta_i = 2\pi i(k+\Delta)/n+\theta$, see (6.6.1).

a) Expanding in a Taylor series for $g(a)=E_1(I)$ yields

$$E_1^2(I) = g^2(0)+2ag(0)\dot{g}(0)+a^2(\dot{g}^2(0)+g(0)\ddot{g}(0))+O(a^3) \tag{I.10}$$

$$\rightarrow \quad E_1^2(I) = a^2\left[\left.\frac{dE_1(I)}{da}\right|_{a=0}\right]^2 + O(a^3)$$

but,

$$E_1(I)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n} c_i \sum_{l=1}^{M+1} a_{M+1}(l)P_{1i}^{x}(l) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i \sum_{l=1}^{M+1} a_{M+1}(l)P_{1i}^{y}(l) \tag{I.11}$$

where

$$P_{1i}^{x}(l) = P_1\{R_{(M+1)i}^{x}=l\} = \binom{M}{l-1}\int f(x-aA\cos\beta_i)F^{l-1}(x)[1-F(x)]^{M+1-l}\,dx \tag{I.12}$$

$$P_{1i}^{y}(l) = P_1\{R_{(M+1)i}^{y}=l\} = \binom{M}{l-1}\int f(x-aA\sin\beta_i)F^{l-1}(x)[1-F(x)]^{M+1-l}\,dx \tag{I.13}$$

Thus,

$$\left.\frac{dP_{1i}^{x}(l)}{da}\right|_{a=0} = \frac{Aa(l,f)\cos\beta_i}{M+1}\ , \quad \left.\frac{dP_{1i}^{y}(l)}{da}\right|_{a=0} = \frac{Aa(l,f)\sin\beta_i}{M+1} \tag{I.14}$$

where $a(l, f)$ is given by Eq.(F.8) of App. F and is assumed finite (sufficient condition is f to be strongly unimodal). Differentiating (I.11) and substituting (I.14) yields

$$\left.\frac{dE_1(I)}{da}\right|_{a=0} = \frac{AG(f)}{\sqrt{n}}\sum_{i=1}^{n}\cos(\beta_i-\alpha_i)\ ,\ G(f) \triangleq \frac{1}{M+1}\sum_{l=1}^{M+1} a_{M+1}(l)a(l,f) \tag{I.15}$$

In a similar manner,

$$\left.\frac{dE_1(Q)}{da}\right|_{a=0} = \frac{AG(f)}{\sqrt{n}}\sum_{i=1}^{n}\sin(\beta_i-\alpha_i) \tag{I.16}$$

Combining (I.10) with (I.15 - I.16) and $\beta_i-\alpha_i=\delta_i$ of (I.1),

$$E_1^2(I)+E_1^2(Q) = na^2A^2G^2(f)\left|\frac{1}{n}\sum_{i=1}^{n}\exp(j2\pi i\Delta/n)\right|^2 + O(na^3) \tag{I.17}$$

Notice that θ has canceled. Due to the i.i.d. assumption and from the mean value

theorem,

$$Var_1(I) = \frac{1}{n}\sum_{i=1}^{n} Var_1(I_i) = Var_0(I) + \frac{a}{n}\sum_{i=1}^{n} \frac{dVar_1(I_i)}{da}\bigg|_{a' \le a} \tag{I.18}$$

$$\le Var_0(I) + a \max_i \{\frac{dVar_1(I_i)}{da}\bigg|_{a' \le a}\}$$

Hence, a sufficient condition for $\lim_{a\to 0} Var_1(I) = Var_0(I)$, is $\lim_{a\to 0} dVar_1(I_i)/da < \infty$, $\forall i$.

Since

$$Var(I_i) = c_i^2 Var(U_i) + s_i^2 Var(V_i) = Var(U_i) = Var(V_i) = Var(Q_i)$$

and

$$Var_1(U_i) = \sum_{l=1}^{M+1} a^2(l) P_{1i}^x(l) - \sum_{l=1}^{M+1}\sum_{k=1}^{M+1} a(l)a(k) P_{1i}^x(l) P_{1i}^x(k) \tag{I.19}$$

with a similar expression for $Var_1(V_i)$, it is seen from (I.14) that the derivatives of $Var_1(I_i)$ and $Var_1(Q_i)$ are proportional to $a(l,f)$, hence bounded, thus

$$\lim_{a\to 0} Var_1(I) = \lim_{a\to 0} Var_1(Q) = E_0^2(V_i) = E_0^2(U_i) = \frac{1}{M+1}\sum_{l=1}^{M+1} a_{M+1}^2(l) \tag{I.20}$$

Combining (I.17) and (I.20)

$$\lim_{H_1\to H_0} \frac{E_1^2(I)}{Var_1(I)} + \frac{E_1^2(Q)}{Var_1(Q)} = \frac{na^2A^2G^2(f)G(\Delta)}{E_0^2(U_i)} \tag{I.21}$$

b) From (I.15 - I.16), the first term in the expansion is

$$\frac{IE_1(I) + QE_1(Q)}{aAG(f)/\sqrt{n}} = I\sum_{i=1}^{n}\cos\delta_i + Q\sum_{i=1}^{n}\sin\delta_i =$$

$$= \cos\theta[I\sum_{i=1}^{n}\cos\gamma_i + Q\sum_{i=1}^{n}\sin\gamma_i] + \sin\theta[-I\sum_{i=1}^{n}\sin\gamma_i + Q\sum_{i=1}^{n}\cos\gamma_i] =$$

$$= \sqrt{C^2+S^2}\cos(\theta - \tan^{-1}\frac{S}{C})$$

where C is the term multiplying $\cos\theta$, S is the term multiplying $\sin\theta$, and $\gamma_i = 2\pi i\,\Delta/n$.

It is easy to show

$$C^2+S^2 = (I^2+Q^2)\left|\sum_{i=1}^{n}\exp(j\gamma_i)\right|^2 \underset{n\to\infty}{\longrightarrow} (I^2+Q^2)n^2G(\Delta)$$

Thus,

$$\lim_{H_1 \to H_0} \frac{IE_1(I)}{Var_1(I)} + \frac{QE_1(Q)}{Var_1(Q)} = \frac{\sqrt{n}\, aG(f)G^{1/2}(\Delta)}{E_0{}^2(U_i)} \cos(\theta - \tan^{-1}\frac{S}{C}) \qquad \text{(I.22)}$$

c) We neeed to show that asymptotically I and Q are uncorrelated. Starting from (I.8) and invoking the appropriate independence assumptions, it can be shown

$$E_1(IQ) - E_1(I)E_1(Q) = \frac{1}{n}\sum_{i,j} s_i\, c_j\, (E_1(V_i\, V_j) - E_1(V_i)E_1(V_j)) \qquad \text{(I.23)}$$

$$- \frac{1}{n}\sum_{i,j} c_i\, s_j\, (E_1(U_i\, U_j) - E_1(U_i)E_1(U_j)) = \frac{1}{n}\sum_{i=1}^{n} s_i\, c_i\, (Var_1(V_i) - Var_1(U_i))$$

Expanding,

$$= \frac{1}{n}\sum_{i=1}^{n} s_i\, c_i\, (Var_0(V_i) - Var_0(U_i)) + \frac{a}{n}\sum_{i=1}^{n} s_i\, c_i \left.\frac{d\,(Var_0(V_i) - Var_0(U_i))}{da}\right|_{0 \le a' \le a}$$

$$\le a \max_i \left| \frac{d\,(Var_0(V_i) - Var_0(U_i))}{da} \right| \Bigg|_{0 \le a' \le a} = O(a)$$

Thus, as the variances are O(1),

$$\lim_{H_1 \to H_0} \frac{E_1(IQ) - E_1(I)E_1(Q)}{\sqrt{Var_1(I)\, Var_1(Q)}} = 0$$

Notice that cancellation of the zeroth order of the covariance is due to the rotation transformation, which results in (I.23). In fact, it can be shown that the derivative of the difference is proportional to $\sum_{l=1}^{M+1} a_{M+1}^2(l)\, a(l,\, f)$. Thus, with the additional assumption of symmetric f this term also vanishes, and the correlation coefficient is even $O(a^2)$

Appendix J

Monotonicity of the error probability in digital noncoherent communication with respect to the SNR [1]

In Chap. 8, it was shown that the generalized SNR $P(l,\ f\ )$ possess a minimax property, in the particular class of AWNGN processes and in the class of decision rules based on either limiter-correlators or on generalized M-estimators. To show a minimax property on the error probability, it suffices to show that the error probability of the optimal M-ary decision rules for the AWGN channel is a monotonic function of the SNR. While this seems obvious and is evident from all the graphic exhibits in the literature, we have not find any general proof for arbitrary signaling schemes and noncoherent reception. In the following, we demonstrate it for the important case of orthogonal signals. We denote by $X=\epsilon/N_0$ the SNR.

The probability of correct decision is given by, cf. [65],

$$P_c = 1 - \frac{1}{M}\sum_{i=1}^{M}(-1)^i\binom{M}{i}e^{(1-i)X/i} \tag{J.1}$$

Differentiating, we have to show the following for all $X>0$ and $M\geq 2$

$$f(x) = \sum_{i=1}^{M}(-1)^i\binom{M}{i}\frac{i-1}{i}e^{X/i} \geq 0 \tag{J.2}$$

Expanding in a Taylor series at $X=0$, it suffices to show that the coefficients a_k are positive for all $k\geq 0$, where

$$a_k = \frac{\partial^k f(X)}{\partial X^k}\Bigg|_{X=0} = \sum_{i=1}^{M}(-1)^i\binom{M}{i}\frac{i-1}{i^{k+1}} \tag{J.3}$$

Define a generating function $P_k(t)$ to be given by the same expression for a_k, but the summands are multiplied by t^i; thus $P_k(0)=0$ and $P_k(1)=a_k$. We will prove that $P_k(t)$ is positive for all $t>0$, by mathematical induction on k. a) Assume $P_{k-1}(t)$ is

[1] The initial part of the proof was suggested by Avner Dor.

positive, than it is easy to see that

$$P_k(t) = \int_0^t \frac{P_{k-1}(y)}{y} dy \tag{J.4}$$

and is positive by the assumption. b) To show that $P_0(t)$ is positive, let $P_0(t) = g(t) - h(t)$, where

$$g(t) = \sum_{i=1}^{M} (-1)^i \binom{M}{i} t^i = (1-t)^M - 1 = M \int_0^t (1-y)^{M-1} dy \tag{J.5}$$

(the middle equality follows from the binomial expansion), and

$$h(t) = \sum_{i=1}^{M} (-1)^i \binom{M}{i} \frac{t^i}{i} \tag{J.6}$$

It ie easy to see that $t\dot{h}(t) = (1-t)^M - 1$, thus, as $h(0)=0$,

$$h(t) = \int_0^t \frac{(1-y)^M - 1}{y} dy \tag{J.7}$$

Combining with (J.5),

$$P_0(t) = g(t) - h(t) = \int_0^t \frac{1 - (1-y)^{M-1}(1-(M+1)y)}{y} dy \tag{J.8}$$

It can be seen that the numerator of the integrand is positive for $0 < y \leq t \leq 1$, thus $P_0(t)$ is clearly positive and the proof is completed.

9. REFERENCES

1. P.J. Huber, "Robust estimation of a location parameter," *Ann. Math. Stat.,* vol. 35, pp. 73-101, 1964.

2. P.J. Huber, "A robust version of the probability ratio test," *Ann. Math. Stat.,* vol. 36, pp. 1753-1758, 1965.

3. P.J. Huber, "Robust Statistics," Wiley & Sons, New York, 1981.

4. R.D. Martin and S.C. Schwartz, "Robust detection of a known signal in nearly Gaussian noise," *IEEE Trans. Inform. Theory,* vol. IT-17, pp. 50-56, January 1971.

5. S.A. Kassam and J.B. Thomas, "Asymptotically robust detection of a known signal in contaminated non-Gaussian noise," *IEEE Trans. Inform. Theory,* vol. IT-22, pp. 22-26, January 1976.

6. A.H. El-Sawy and V.D. VandeLinde, "Robust detection of known signals," *IEEE Trans. Inform. Theory,* vol. IT-23, pp. 722-727, November 1977.

7. H.V. Poor and J.B. Thomas, "Asymptotically robust quantization for detection," *IEEE Trans. Inform. Theory,* vol. IT-24, pp. 222-229, 1978

8. S.A. Kassam, "Locally robust array detectors for random signals," *IEEE Trans. Inform. Theory,* vol. IT-24, pp. 309-316, May 1978.

9. H.V. Poor, M. Mami and J.B. Thomas, "On robust detection of discrete-time stochastic signals," *J. Franklin Inst.,* vol. 309, pp. 29-53, January 1980.

10. S.A. Kassam, G. Moustakides and J.G. Shin, "Robust detection of known signals in asymmetric noise," *IEEE Trans. Inform. Theory,* vol. IT-28, pp. 84-91, 1982.

11. J.G. Shin and S.A. Kassam, "Robust detector for narrowband signals in non-Gaussian noise," *J. Acoust. Soc. Amer.,* vol. 74, pp. 527-533, August 1983.

12. S.A. Kassam, "Detection of narrowband signals: Asymptotic robustness and quantization," *1984 Conference on Information Sciences and Systems,* Princeton University, March 1984.

13. J.V. DiFranco and W.L. Rubin, "Radar detection," Prentice Hall, New Jersey, 1968.

14. H.L. VanTrees, "Detection, estimation and modulation theory; Part III: Radar-Sonar signal processing and Gaussian signals in noise," Wiley & Sons, New York, 1971.

15. C.W. Helstrom, "Statistical theory of signal detection," Second Edition, Pergamon Press, Oxford, 1968.

16. J.B.G. Roberts, "The impact of new VLSI technology on radar signal processing," *Proc. of the International Conference on Radar,* pp. 9-15, Paris, 1984.

17. P. Swerling, "Probability of detection for fluctuating targets," Rand Report RM-1217, March 1954. (Reprinted in *IRE Trans. Inform. Theory,* vol. IT-6, pp. 269-308, April 1960.)

18. H.M. Finn and R.S. Johnson, "Adaptive detection mode with threshold control as function of spatially sampled clutter-level estimates," *RCA Review,* vol. 30, pp. 414-465, September 1968.

19. V.G. Hansen, "Constant false alarm rate processing in search radars," *Proc. of the IEEE 1973 Inter. Radar Conf.,* pp. 325-332, London, 1973.

20. M. Weiss, "Analysis of some modified cell-averaging CFAR processors in multiple target situations," *IEEE Trans. on Aeros. Elect. Syst.,* vol. AES-18, no. 1, pp. 102-114, January 1982.

21. J.D. Gibson and J.L. Melsa, "Introduction to nonparametric detection with applications," Academic Press, New York, 1975.

22. A.A. Winder, "Underwater sound - a review; Part II: Sonar system technology," *IEEE Trans. on Sonics and Ultrasonics,* vol. SU-22, no. 5, pp. 291-332, 1975.

23. A. Wald, "Asymptotically most powerful tests of statistical hypotheses," *Ann. Math. Statist.,* vol. 12, pp. 1-19, March 1941.

24. H. Chernoff, "On the distribution of the likelihood ratio," *Ann. Math. Statist.,* vol. 25, no. 3, pp. 573-578, September 1954.

25. J. Capon, "On the asymptotic relative efficiency of locally optimum detectors," *IRE Trans. Inform. Theory,* vol. IT-7, pp. 67-72, April 1961.

26. D. Middleton, "An introduction to statistical communication theory," McGraw-Hill, New York, 1960.

27. D.A.S. Fraser, "Nonparametric methods in statistics," Academic Press, New York, 1951.

28. T.S. Ferguson, "Mathematical statistics," Academic Press, New York, 1967.

29. A.B. Martinez, "Asymptotic performance of detectors with non-zero input SNR," *Proc. 20th Annual Allerton Conf. on Commun., Control and Computing,* University of Illinois, Urbana-Champaign, IL, pp. 749-758, 1982.

30. D.L. Michalsky, G.L. Wise and H.V. Poor, "A relative efficiency study of some popular detectors," *J. of Franklin Inst.,* vol. 313, pp. 135-148, March 1982.

31. A.D. Spaulding, "Locally optimum and suboptimum detector performance in non-Gaussian noise," *Proc. of IEEE ICC 1982,* vol. 7, pp. 2H.2.1-7, June 1982.

32. J.B. Thomas, "Nonparametric detection," *Proc. IEEE,* vol. 58, no. 5, pp. 623-631, 1970.

33. P.J. Bickel, "On some robust estimates of location," *Ann. Math. Statist.,* 36, pp. 847-858, 1965.

34. F.R. Hampel, "The influence curve and its role in robust estimation," *J. Amer. Statist. Ass.,* 62, pp. 1179-1186, 1974.

35. D.F. Andrews, et al., "Robust estimation of location, survey and advances," Princeton, New Jersey, Princeton University Press, 1972.

36. S.M. Kendal and A. Stuart, "The advanced theory of statistics," vol. 1, Fourth Edition, Charles Griffin & Company, London, 1977.

37. W. Feller, "An introduction to probability theory and its applications," vol. II, Wiley & Sons, New York, 1966.

38. G. Szego, "Orthogonal Polynomials," New York, 1939.

39. A.V. Aho, J.E. Hopcroft and J.D. Ullman, "The design and analysis of computer algorithms," Addison-Wesley, Reading, MA, 1974.

40. V.G. Hansen, "Detection performance of the narrow-band Wilcoxon detector against Gaussian noise," *IEEE Trans. on Inform. Theory,* vol. IT-18, pp. 664-667, September 1972.

41. P. Billingsley, "Probability and measure," Wiley & Sons, New York, 1979.

42. C.A. Field and F.R. Hampel; "Small-sample asymptotic distribution of M-estimates of location," *Biometrika,* vol. 69, no. 1, pp. 29-46, 1982.

43. H.E. Daniels, "Saddlepoint approximations in statistics," *Ann. Math. Statist.,* vol. 25, pp. 631-650, 1954.

44. R.L. Mitchell, "Importance sampling applied to simulation of false alarm statistics," *IEEE Trans. on Aeros. Elect. Syst.,* vol. AES-17, no. 1, pp. 15-24, January 1981.

45. M. Weiss and S.C. Schwartz, "On optimal minimax jamming and detection of radar signals," *IEEE Trans. on Aeros. and Elect. Syst.,* Vol. AES-21, no. 3, pp. 385-393, May 1985.

46. L.A. Jaeckel, "Robust estimates of location: Symmetric and asymmetric contamination," *Ann. Math. Statist.,* 42, pp. 1020-1034, 1971.

47. S.M. Stigler, "Linear functions of order statistics with smooth weight functions," *Ann. Statistics,* 2, pp. 676-693, 1974.

48. M. Weiss and S.C. Schwartz, "Robust detection of coherent radar signals in nearly Gaussian noise," Proc. of the IEEE 1985 International Radar Conference, pp. 297-

302, Washington, May 1985.

49. J. Hájek and Z. Šidák, "Theory of rank tests," Academic Press, New York, 1967.

50. J. Hájek, "Nonparametric Statistics," Holden-Day, San Francisco, 1969.

51. A.D. Whalen, "Detection of signals in noise," Academic Press, New York, 1971.

52. G.M. Dillard and C.E. Antoniak, "A practical distribution-free detection procedure for multiple-range-bin radars," *IEEE Trans. on Aerospace and Electronic Systems,* Vol. AES-6, pp. 629-635, Sept. 1970.

53. V.G. Hansen and B.A. Olsen, "Nonparametric radar extraction using a generalized sign test," *IEEE Trans. on Aerospace and Electronic Systems,* Vol. AES-7, pp. 942-950, Sept. 1971.

54. G.W. Zeoli and T.S. Fong, "Performance of a two-sample Mann-Whitney nonparametric detector in a radar application," *IEEE Trans. on Aerospace and Electronic Systems,* Vol. AES-7, pp. 951-959, Sept. 1971.

55. E.K. Al-Hussaini, F.M. Badran and L.F. Turner, "Modified savage and modified rank squared nonparametric detectors," *IEEE Trans. on Aerospace and Electronic Systems,* Vol. AES-14, pp. 242-250, March 1978.

56. S.A. Kassam, "A bibliography on nonparametric detection," *IEEE Trans. on Information Theory,* Vol. IT-26, pp. 595-602, Sept. 1980.

57. E.A. Feustel and L.D. Davisson, "The asymptotic relative efficiency of mixed statistical tests," *IEEE Trans. on Information Theory,* Vol. IT-13, pp. 247-255, April 1967.

58. E.A. Feustel and L.D. Davisson, "On the efficacy of mixed locally most powerful one- and two-sample rank tests," *IEEE Trans. on Information Theory,* Vol. IT-14, pp. 776-778, Sept. 1968.

59. H. Chernoff and I.R. Savage, "Asymptotic normality of certain nonparametric test statistics," *Annals of Math. Stat.,* Vol. 29, pp. 972-994, 1958.

60. L. LeCam, "Locally asymptotically normal families of distributions," Univ. of California Publ. in Stat. 3, pp. 37-98, 1960.

61. J. Hájek, "Asymptotically most powerful rank-order tests," *Annals of Math. Stat.,* Vol. 33, pp. 1124-1147, 1962.

62. W. Hoeffding, "Optimum nonparametric tests," *Proc. 2nd Berkeley Symp. on Math. Stat. and Prob.,* Berkeley, California, pp. 83-92, 1951.

63. J. Capon, "Asymptotic efficiency of certain locally most powerful rank tests," *Annals of Math. Stat.,* Vol. 32, pp. 88-100, 1961.

64. K.L. Chung, "A Course in Probability Theory", Academic Press, New York, 1974.

65. A.J. Viterbi, "Principles of Coherent Communication", McGraw-Hill, New York, 1966.

66. R.G. Gallager, "Information Theory and Reliable Communication", John Wiley and Sons, New York, 1968.

67. A.J. Viterbi and J.K. Omura, "Principles of Digital Communication and Coding", McGraw-Hill, New York, 1979.

68. J.G. Proakis, "Digital Communications", McGraw-Hill, New York, 1983.

69. R.D. Martin and S.C. Schwartz, "On mixture, quasi-mixture and nearly normal random processes", *Annals of Math. Stat.,* Vol. 43 No. 3, pp. 948-967, 1972.

70. R.D. Martin, "Robust detection and estimation", Ph.D. Dissertation, Princeton University, 1969.

71. M. Rosenblatt, "Random Processes", Oxford Univ. Press, 1962.

72. N.C. Beaulieu and C. Leung, "On the performance of three suboptimum detection scheme for binary signaling", *IEEE Trans. on Communications,* Vol. Com-33 No.3, pp.241-245, March 1985.

73. H. Rieder, "A robust asymptotic testing model", *Ann. Statist.,* Vol. 6, pp. 1080-1094, 1978.

74. S.L. Portnoy, "Robust estimation in dependent situations", *Ann. Statist.,* Vol. 5, pp. 22-43, 1977.

75. H.V. Poor, "Signal detection in the presence of weakly dependent noise - Part II: Robust detection", *IEEE Trans. Inform. Theory,* Vol. IT-28, pp. 744-752, 1982.

76. G.V. Moustakides and J.B. Thomas, "Min-Max detection of weak signals in ϕ–mixing noise", *IEEE Trans. Inform. Theory,* Vol. IT-30, pp. 529-537, May 1984.

77. R.D. Martin, "Robust estimation of signal parameters with dependent data", *Proc. 21st IEEE Conf. on Decision and Control (Orlando, FL, Dec. 8-10, 1982),* pp. 433-436.

78. S. Haykin, "Radar signal processing", *IEEE ASSP Magazine,* Vol. 2 No.2, pp. 2-18, April 1985.

79. M.B. Sirvanky and S.S. Wolff, "Nonparametric detection with autoregressive data", *IEEE Trans. Inform. Theory,* Vol. IT-22, pp. 725-731, November 1976.

OFFICE OF NAVAL RESEARCH
STATISTICS AND PROBABILITY PROGRAM

BASIC DISTRIBUTION LIST
FOR
UNCLASSIFIED TECHNICAL REPORTS

FEBRUARY 1982

	Copies
Statistics and Probability Program (Code 411(SP)) Office of Naval Research Arlington, VA 22217	3
Defense Technical Information Center Cameron Station Alexandria, VA 22314	12
Commanding Officer Office of Naval Research Eastern/Central Regional Office Attn: Director for Science Barnes Building 495 Summer Street Boston, MA 02210	1
Commanding Officer Office of Naval Research Western Regional Office Attn: Dr. Richard Lau 1030 East Green Street Pasadena, CA 91101	1
U. S. ONR Liaison Office - Far East Attn: Scientific Director APO San Francisco 96503	1
Applied Mathematics Laboratory David Taylor Naval Ship Research and Development Center Attn: Mr. G. H. Gleissner Bethesda, Maryland 20084	1
Commandant of the Marine Coprs (Code AX) Attn: Dr. A. L. Slafkosky Scientific Advisor Washington, DC 20380	1
Navy Library National Space Technology Laboratory Attn: Navy Librarian Bay St. Louis, MS 39522	1
U. S. Army Research Office P.O. Box 12211 Attn: Dr. J. Chandra Research Triangle Park, NC 27706	1
Director National Security Agency Attn: R51, Dr. Maar Fort Meade, MD 20755	1
ATAA-SL, Library U.S. Army TRADOC Systems Analysis Activity Department of the Army White Sands Missile Range, NM 88002	1
ARI Field Unit-USAREUR Attn: Library c/o ODCSPER HQ USAEREUR & 7th Army APO New York 09403	1
Library, Code 1424 Naval Postgraduate School Monterey, CA 93940	1
Technical Information Division Naval Research Laboratory Washington, DC 20375	1
OASD (I&L), Pentagon Attn: Mr. Charles S. Smith Washington, DC 20301	1

Copies

Director
AMSAA
Attn: DRXSY-MP, H. Cohen
Aberdeen Proving Ground, MD 21005 1

Dr. Gerhard Heiche
Naval Air Systems Command (NAIR 03)
Jefferson Plaza No. 1
Arlington, VA 20360 1

Dr. Barbara Bailar
Associate Director, Statistical Standards
Bureau of Census
Washington, DC 20233 1

Leon Slavin
Naval Sea Systems Command (NSEA 05H)
Crystal Mall #4, Rm. 129
Washington, DC 20036 1

B. E. Clark
RR #2, Box 647-B
Graham, NC 27253 1

Naval Underwater Systems Center
Attn: Dr. Derrill J. Bordelon
Code 601
Newport, Rhode Island 02840 1

Naval Coastal Systems Center
Code 741
Attn: Mr. C. M. Bennett
Panama City, FL 32401 1

Naval Electronic Systems Command (NELEX 612)
Attn: John Schuster
National Center No. 1
Arlington, VA 20360 1

Defense Logistics Studies Information Exchange
Army Logistics Management Center
Attn: Mr. J. Dowling
Fort Lee, VA 23801 1

Copies

Reliability Analysis Center (RAC)
RADC/RBRAC
Attn: I. L. Krulac
Data Coordinator/
Government Programs
Griffiss AFB, New York 13441 1

Technical Library
Naval Ordnance Station
Indian Head, MD 20640 1

Library
Naval Ocean Systems Center
San Diego, CA 92152 1

Technical Library
Bureau of Naval Personnel
Department of the Navy
Washington, DC 20370 1

Mr. Dan Leonard
Code 8105
Naval Ocean Systems Center
San Diego, CA 92152 1

Dr. Alan F. Petty
Code 7930
Naval Research Laboratory
Washington, DC 20375 1

Dr. M. J. Fischer
Defense Communications Agency
Defense Communications Engineering Center
1860 Wiehle Avenue
Reston, VA 22090 1

Mr. Jim Gates
Code 9211
Fleet Material Support Office
U. S. Navy Supply Center
Mechanicsburg, PA 17055 1

Mr. Ted Tupper
Code M-311C
Military Sealift Command
Department of the Navy
Washington, DC 20390 1

	Copies	Copies
Mr. F. R. Del Priori Code 224 Operational Test and Evaluation Force (OPTEVFOR) Norfolk, VA 23511	1	